

Optical System Design

Robert E. Fischer
CEO, OPTICS 1, Incorporated

Biljana Tadic-Galeb
Panavision

Paul R. Yoder
Consultant

With contributions by
Ranko Galeb
Bernard C.Kress, Ph.D.
Stephen C. McClain, Ph.D.
Tom Baur
Rick Plympton
Bob Wiederhold
Alastair J. Grant

Second Edition



New York · Chicago · San Francisco · Lisbon · London
Madrid · Mexico City · Milan · New Delhi · San Juan
Seoul · Singapore · Sydney · Toronto

Copyright © 2008 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-159358-6

The material in this eBook also appears in the print version of this title: 0-07-147248-7.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/0071472487

CONTENTS

Preface	xiii	
Acknowledgments	xv	
Chapter 1.	Basic Optics and Optical System Specifications	1
	The Purpose of an Imaging Optical System	1
	How to Specify Your Optical System: Basic Parameters	4
	Basic Definition of Terms	11
	Useful First-Order Relationships	15
Chapter 2.	Stops and Pupils and Other Basic Principles	29
	The Role of the Aperture Stop	29
	Entrance and Exit Pupils	31
	Vignetting	32
Chapter 3.	Diffraction, Aberrations, and Image Quality	35
	What Image Quality Is All About	35
	What Are Geometrical Aberrations and Where Do They Come From?	36
	What Is Diffraction?	40
	Diffraction-Limited Performance	43
	Derivation of System Specifications	45
Chapter 4.	The Concept of Optical Path Difference	49
	Optical Path Difference (OPD) and the Rayleigh Criteria	49
	Peak-to-Valley and RMS Wavefront Error	52
	The Wave Aberration Polynomial	55
	Depth of Focus	56
Chapter 5.	Review of Specific Geometrical Aberrations and How to Get Rid of Them	59
	Spherical Aberration	60
	Coma	72
	Astigmatism	75

Field Curvature and the Role of Field Lenses	78
Distortion	85
Axial Color	89
Lateral Color	90
Parametric Analysis of Aberrations Introduced by Plane Parallel Plates	91
Chapter 6. Glass Selection (Including Plastics)	95
Material Properties Overview	95
The Glass Map and Partial Dispersion	96
Parametric Examples of Glass Selection	102
How to Select Glass	106
Plastic Optical Materials	109
A Visual Aid to Glass Selection	111
Chapter 7. Spherical and Aspheric Surfaces	115
Definition of an Aspheric Surface	115
Conic Surfaces	117
Application of Aspheric Surfaces in Reflective and Refractive Systems	119
Guidelines in the Use of Aspheric Surfaces	124
Specification of Aspheric Surfaces	126
Chapter 8. Design Forms	129
Introduction	129
System Configurations for Refractive Systems	131
System Configurations for Reflective Systems	138
Reflective Systems, Relative Merits	144
Refractive Systems, Relative Merits	146
Mirrors and Prisms	147
Design of Visual Systems	155
Chapter 9. The Optical Design Process	167
What Do We Do When We Optimize a Lens System?	168
How Does the Designer Approach the Optical Design Task?	171
Sample Lens Design Problem	176

Chapter 10.	Computer Performance Evaluation	179
What Is Meant by Performance Evaluation	179	
What Is Resolution?	180	
Ray Trace Curves	181	
Spot Diagrams	187	
Optical Path Difference	189	
Encircled Energy	189	
MTF	191	
Chapter 11.	Gaussian Beam Imagery	199
Beam Waist and Beam Divergence	201	
Collimation of Laser Beams	203	
Propagation of Gaussian Beams and Focusing into a Small Spot	204	
Truncation of a Gaussian Beam	205	
Application of Gaussian Beam Optics in Laser Systems	208	
F-θ Lenses in Laser Scanners	211	
Chapter 12.	Basics of Thermal Infrared Imaging in the 3- to 5- and 8- to 12- μm Spectral Bands (Plus UV Optics)	213
The Basics of Thermal Infrared Imaging	213	
The Dewar, Cold Stop, and Cold Shield	217	
Cold Stop Efficiency	219	
Scanning Methods	222	
IR Materials	229	
Reduced Aberrations with IR Materials	236	
Image Anomalies	239	
Athermalization	246	
System Design Examples	250	
Optical Systems for the UV	255	
Chapter 13.	Diffractive Optics	259
Introduction	259	
The Many Faces of Diffractive Optics	262	
What Design and Modeling Tools Should I Use?	277	
How Are Diffractives Fabricated?	287	
Where Are Diffractives Used?	308	
References	318	

Chapter 14.	Design of Illumination Systems	321
	Introduction	321
	Köhler and Abbe Illumination	322
	Optical Invariant and Etendue	324
	Other Types of Illumination Systems	329
Chapter 15.	Performance Evaluation and Optical Testing	333
	Testing with the Standard 1951 U.S. Air Force Target	333
	The Modulation Transfer Function	337
	Interferometry	340
	Other Tests	344
Chapter 16.	Tolerancing and Productibility	347
	Introduction	347
	What Are Testplates and Why Are They Important?	348
	How to Tolerance an Optical System	353
	How Image Degradations from Different Tolerances Are Summed	356
	Forms of Tolerances	359
	Adjusting Parameters	364
	Typical Tolerances for Various Cost Models	366
	Example of Tolerance Analysis	367
	Surface Irregularities	374
	How Does Correlation Relate to Performance?	376
	Effect to Spot Diameter	377
	Effect to MTF: The Optical Quality Factor	379
	Beam Diameter and Surface Irregularity	383
	The Final Results	384
Chapter 17.	Optomechanical Design	389
	Environmental Considerations	389
	Applicable Design Guidelines	393
	Environmental Testing Methods	393
	Mechanical Parameters and Properties	393
	Typical Mechanical Property Values for Selected Materials	394
	Structural Design	396
	Vibration, Self-Weight Deflection, and Fundamental Frequency	398
	Shock	400

Rigid Housing Configurations	400
Modular Construction	401
Support Structure Configurations	405
Establishing Axial and Lateral Preload Requirements	414
Spherical and Crowned Lens Rims	415
Interfaces for Other Optical Components	416
Individual Lens Mounting Techniques	419
Surface Contact Interface Shapes	426
Mounting Windows, Shells, and Domes	429
Stress Consequences of Axial Preload	434
Temperature Effects on Axial Preload	436
Radial Stresses and Their Variations with Temperature	439
Bending Effects in Rotationally Symmetric Optics	439
Multiple-Component Lens Assemblies	441
Incorporating Prisms into the Design	452
Mirror Mountings	459
Mechanical Athermalization Techniques	467
References	476
Chapter 18. Optical Manufacturing Considerations	479
Material	480
Manufacturing	485
Special Fabrication Considerations	492
Relative Manufacturing Cost	502
Sourcing Considerations	502
Conclusion	504
Chapter 19. Polarization Issues in Optical Design	507
Introduction	507
Introduction to Polarization	508
The Mathematical Description of Polarized Light	513
Some Polarization Phenomena	523
Polarization Control Nuts and Bolts	535
Polarization Analysis of an Optical System	555
Minimizing Polarization Problems in Optical Design	559
Polarization as a Tool in Optical System Design	560
Summary	565
Bibliography	567

Chapter 20.	Optical Thin Films	569
	Introduction	569
	Designing Optical Coatings	570
	Various Categories of Optical Coatings	571
	Optical Coating Process	578
	Coating Performance Versus Number of Layers	582
	Specifying Coating Requirements	583
	Relationship Between Production Cost, Tolerances, and Quality	584
	Bibliography	585
Chapter 21.	Hardware Design Issues	587
	Off-the-Shelf Optics	587
	How to Effectively Work with Off-the-Shelf Optics	589
	Working with Off-the-Shelf Singlets and Doublets	590
	Example of Lens Used at Conjugates Different from What It Was Designed	591
	Pupil Matching	594
	Development of a Lab Mockup Using Off-the-Shelf Optics	595
	Stray Light Control	595
	Optomechanical Design	600
Chapter 22.	Lens Design Optimization Case Studies	603
	Error Function Construction	603
	Achromatic Doublet Lens Design	605
	Double Gauss Lens Design	610
	Digital Camera Lens	632
	Binocular Design	642
	Parametric Design Study of Simple Lenses Using Advanced Manufacturing Methods	646
	Design Data for Double Gauss	655
Chapter 23.	Optical Sensor Systems Modeling and Analysis	659
	Introduction	659
	Image Formation	660
	Detector Arrays	663
	Optical System Noise Characteristics	669

Color Sensors	691
Electronic Correction	696
Camera Connectivity	697
Bibliography	701
Chapter 24. Stray Light and Optical Scattering	703
Introduction	703
Stray Light Scatter Sources	703
Types of Scatter	711
Modeling and Analysis Techniques	713
Veiling Glare	715
Cleanliness	716
Suppression Techniques	717
Bright Field and Dark Field	731
How to Avoid Unwanted Stray Light	736
Bibliography	737
Chapter 25. Bloopers and Blunders in Optics	739
Distortion in a 1:1 Imaging Lens	739
Zoom Periscope	740
Sign of Distortion	742
Lens Elements That Are Not Necessary	744
Pupil Problems	744
Not Enough Light	745
Athermalization Using Teflon	746
Athermalization Specifications	746
Bad Glass Choice	747
Elements in Backward	747
Insufficient Sampling of Fields of View or Aperture	748
Images Upside Down or Rotated	749
The Hubble Telescope Null Lens Problem	750
Wrong Glass Type in a Precision Lens System	755
Single Use Camera with a Diffractive Achromat	755
Wrong Image Handedness	756
Cemented Triplet as Part of an Imaging System	757
Total Internal Reflection in a Cube Beamsplitter	758
Diffractive Optics Issues	760
Case of the Miscoated Mangin	763
Telescopes and Polarization	765

Chapter 26.	Rule of Thumb and Hints	767
	General Optical Design Topics	767
	Optomechanical Topics	770
	Diffractive Optics	772
	Glossary	775
	Bibliography	785
	Index	787

PREFACE

The design of imaging optical systems is an engineering discipline which has been practiced and written about for many years. In many ways, optical design is both a science and an art, and for this reason it is a technology that can cause problems if it is not done properly. Furthermore, most books on the subject tend to be complex and difficult to follow and to understand. With this book, we hope to bring the understanding of our discipline to everyone.

We are all aware of cameras, binoculars, and other optical systems and instruments. In the past several years, the field of optics and photonics has seen a tremendous surge in both technology and in applications. This is fueled by a closer association with electronics in devices such as digital cameras, enhanced machine vision systems, MEMS and microoptical systems for telecommunications and other related applications, many of which have yet to be invented.

With this surge in the applications of optics, the educational process of training experienced optical designers and engineers becomes extremely important if not critical.

We realize that it is difficult to be an expert in everything. We also realize that in addition to optical design which is the core of the book, important topics including optical manufacturing, polarization, and optical coatings are important subjects that need to be covered in this book, and the first edition included these topics. With this new second edition, other critical technologies including optomechanical design, systems modeling and analysis, and stray light suppression are now included. Further, completely revised chapters on diffractive optics and polarization are also included.

We are honored to have contributed chapters written by experts in their fields: Paul Yoder on optomechanical design, Rick Plympton and Bob Wiederhold on optical manufacturing, Steve McClain and Tom Baur on polarization in optical systems, Ranko Galeb on thin films and optical coatings, Bernard Kress on diffractive optics, and Alastair J. Grant on systems modeling and analysis, as well as stray light suppression.

The ultimate goal of this book is to teach optical design and engineering in a fully unintimidating way using clear and easy to understand graphics and explanations. Many authors feel an obligation to

include complex mathematical derivations. We have taken a very different approach. We will make this book clear and easy to understand with the goal that you will learn the subject matter with a combination of complete graphics, easy to follow explanations, and just enough math to be useful, but not too much math to make the book hard to follow or difficult to understand.

This book *Optical System Design* is largely based on the firm foundation of the short course by the same title taught by Bob Fischer to over thousands of students over the past 20 years. The course has been honed, polished, and expanded over the years. It is available on CD ROM and videotape, and finally, via this book. Typical comments have been:

"This course was just what I had hoped it would be. It condensed the vast optical world into the key elements necessary for a broad understanding of the subject and an excellent foundation for future study. Good job!"

"Excellent presentation! This is an invaluable course for those who are engaged in optical systems efforts and have a minimum training in optics."

*"A fast paced, well-prepared study, presented by a hands-on instructor."
"I learned what I came to learn, thanks."*

"Excellent! Wonderful presentation and technique. Material was well covered."

"Excellent in explaining and answering questions. Very useful rules of thumb, great presentation. Thanks!"

"Very professional and excellent presentation."

This book is for everyone from program managers to seasoned optical designers and engineers, mechanical engineers, electrical engineers, and others. You will find that it is like reading *Gulliver's Travels*. We all read *Gulliver's Travels* in elementary school, some of us again in high school, and some scholars wrote their Ph.D theses on the book. *Gulliver's Travels* can be read at multiple levels, just like this book.

ROBERT E. FISCHER

ACKNOWLEDGMENTS

Welcome to the second edition of *Optical System Design*. When first published in the year 2000 the goal was to create a book in the field of optical system design and engineering that was clear and easy to understand, and at the same time highly useful to the reader. Far too many books are filled with complex mathematics and other “mumbo jumbo” that often makes it difficult to find what you are really looking for. The overriding goal was thus to create a book that was both extremely useful as well as readable, and all indications are that this goal was met.

I recall talking to a professor at a leading university in optics just after the first edition was published who not only liked the book, but he told me that he was reading the book *backwards*, that is to say from the last chapter forward. “Why” I asked with a puzzled look? Because he really liked the “Bloopers and Blunders in Optics” chapter and that was the last chapter in the book!

I am pleased to welcome four new contributing authors for the second edition. Bernard Kress has a new and fully revised chapter on “Diffractive Optics” and Steve McClain along with Tom Baur have a new and fully revised chapter on “Polarization in Optical Systems.” In addition to the above, I am honored to welcome to the book a new chapter “Optomechanical Design” by Paul Yoder. Optics and mechanics are both critical technologies that truly go hand-in-hand to make for a complete system. We must never underestimate the importance of mechanics to an optical system. Paul Yoder is one of those true icons in the field! Thanks Paul!

Also welcome to Alastair Grant who has contributed chapters on “Optical sensor systems Modeling and Analysis” and also “Stray Light and Optical scattering.”

Finally, a deep and sincere thanks is due to my wife Emilia who so graciously supports my many extracurricular activities in optics. Also thanks to Antonia Petruse Surd, my administrative assistant, who played a major role in coordinating the many inputs from authors.

ROBERT E. FISCHER

OPTICAL SYSTEM DESIGN

Basic Optics and Optical System Specifications

This chapter will discuss what a lens or mirror system does and how we specify an optical system. You will find that properly and completely specifying a lens system early in the design cycle is an imperative ingredient required to design a good system.

The Purpose of an Imaging Optical System

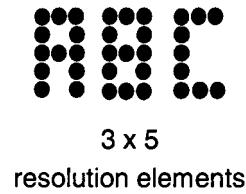
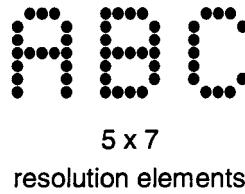
The purpose of virtually all image-forming optical systems is to resolve a specified minimum-sized object over a desired field of view. The *field of view* is expressed as the spatial or angular extent in object space, and the minimum-sized object is the smallest *resolution element* which is required to identify or otherwise understand the image. The word “spatial” as used here simply refers to the linear extent of the field of view in the plane of the object. The field of view can be expressed as an angle or alternatively as a lateral size at a specified distance. For example, the field of view might be expressed as $10^\circ \times 10^\circ$, or alternatively as 350×350 m at a distance of 2 km, both of which mean the same thing.

A good example of a resolution element is the dot pattern in a dot matrix printer. The capital letter E has three horizontal bars, and hence five vertical resolution elements are required to resolve the letter. Horizontally, we would require three resolution elements. Thus, the minimum number of resolution elements required to resolve capital letters is in the vicinity of five vertical by three horizontal. Figure 1.1 is an example of this. Note that the capital letter B and the number 8 cannot be distinguished in a 3×5 matrix, and the 5×7 matrix of dots will do just fine. This applies to telescopes, microscopes, infrared systems, camera lenses, and any other form of image-forming optics. The generally accepted guideline is that approximately three resolution elements or 1.5 line pairs over the object's spatial extent are required to acquire an object. Approximately eight resolution elements or four line pairs are required to recognize the object and 14 resolution elements or seven line pairs are required to identify the object.

There is an important rule of thumb which says that this smallest desired resolution element should be matched in size to the minimum detector element or pixel in a pixelated charged-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS)—type sensor. While not rigorous, this is an excellent guideline to follow for an optimum match between the optics and the sensor. This will become especially clear when we learn about the Nyquist Frequency in Chap. 22, where we show a digital camera design example. In addition, the aperture of the system and transmittance of the optics must be sufficient for the desired sensitivity of the sensor or detector. The detector can be the human eye, a CCD chip or film in your 35-mm camera. If we do not have enough photons to record the imagery, then what good is the imagery?

The preceding parameters relate to the optical system performance. In addition, the design *form* or *configuration* of the optical system must be capable of meeting this required level of performance. For example, most of us will agree that we simply cannot use a single magnifying


Figure 1.1
Illustration of Number of Resolution Elements Required to Resolve or Distinguish Alphanumerics



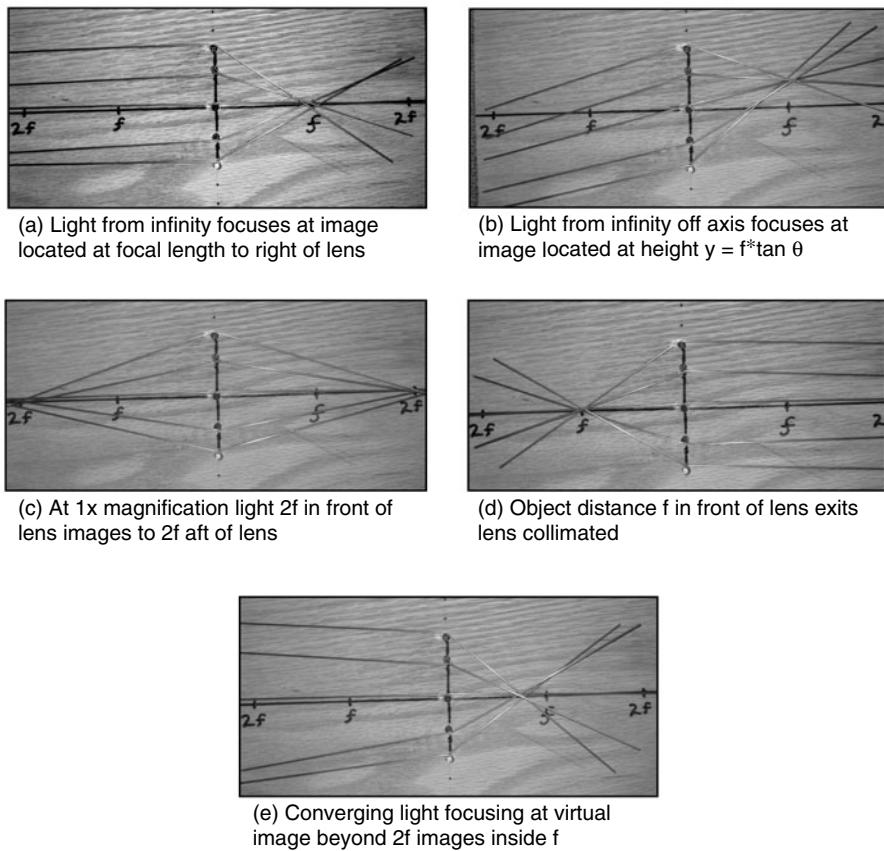
glass element to perform optical microlithography where submicron line-width imagery is required, or even lenses designed for 35-mm or digital photography for that matter. The form or configuration of the system includes the number of lens or mirror elements along with their relative position and shape within the system. We discuss design configurations in Chap. 8 in detail.

Furthermore, we often encounter special requirements, such as cold stop efficiency in infrared systems, scanning systems, and others. These will be addressed later in this book.

Finally, the system design must be producible, meet defined packaging and environmental requirements, weight and cost guidelines, and satisfy other system specifications.

Figure 1.2 shows what we like to call a “wire lens,” and what is shown are photographs of a real construction of a wire lens. If you are new to

Figure 1.2
A Simple “Wire Model” of a Lens Shows how Lenses Work, Where Wires Pivot about Vertical Line Representing Lens



optics, this simple demonstration is an excellent way to learn quickly just how a lens works. Consider (a) where we will show how the wire lens is constructed and how it works. The horizontal centerline that connects the two foci (f) represents the optical axis, and the vertical line represents our lens. If there are no aberrations, light from infinity will have parallel or collimated rays incident onto the lens from the left, and after passing through the lens each of these rays (wires) refract (bend) so as to all pass through the focal point at (f). Each of the five wires is free to pivot about the five black dots on the lens. It is clear that the rays (wires) further from the optical axis bend at a greater angle so that each ray passes through the focus. This is our basic paraxial lens.

Now in (b) we gather the wires at the focus and keeping them together, move the focal point where the wires converge vertically so that they are displaced in the image plane a vertical distance we will call " y ." Note that the rays entering the lens will be parallel or collimated as expected and all be going up to the right by angle θ . This yields the paraxial equation $y = f * \tan \theta$.

Let us now gather the wires together left of the lens and move what is now the object to a position $2f$ in front of the lens. After refraction by the lens the wires will converge at $2f$ behind the lens. It is a well-known property of a lens that an object $2f$ in front of a lens will image $2f$ behind the lens.

In (d) we bring our wires to all pass through (f) in front of the lens, and as expected the exiting wires are parallel, the reverse of (a).

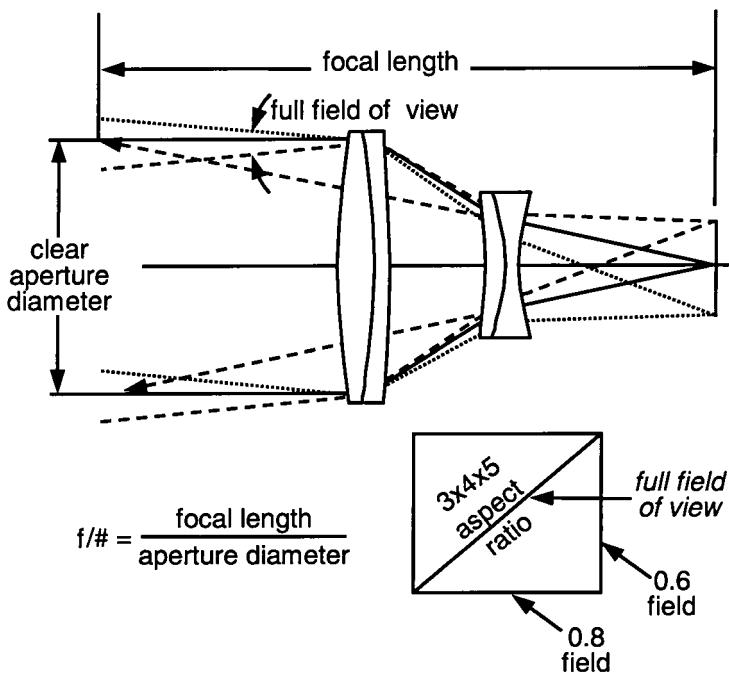
Finally we show in (e) how if the image is inside (f), the wires will be converging entering the lens.

The *wire lens* is very simple to construct, and it is both educational and fun to work with. All you need is piano wire, some small screws, a felt tip Sharpie pen, a piece of wood, and a little bit of time.

How to Specify Your Optical System: Basic Parameters

Consider the lens shown in Fig. 1.3 where light from infinity enters the lens over its *clear aperture diameter*. If we follow the solid ray, we see that it is redirected by each of the lens element groups and components until it comes to focus at the image. If we now extend this ray *backward* from the image toward the front of the system as if it were not bent or

Figure 1.3
Typical Specifications



refracted by the lens groups, it intersects the entering ray at a distance from the image called the *focal length*. The final imaging cone reaching the image at its center is defined by its *f/number* or *f/#*, where

$$f/\text{number} = \frac{\text{focal length}}{\text{clear aperture diameter}}$$

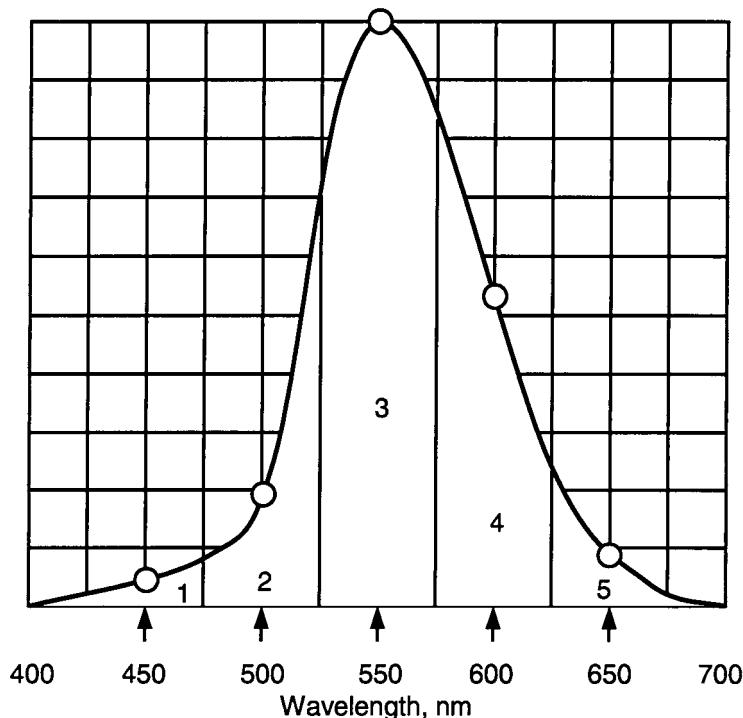
You may come across two other similar terms, *effective focal length* and *equivalent focal length*, both of which are often abbreviated EFL. The effective focal length is simply the focal length of a lens or a group of lenses. Equivalent focal length is very much the same; it is the overall focal length of a group of lens elements, some or all of which may be separated from one another.

The lens is used over a *full field of view*, which is expressed as an angle, or alternatively as a linear distance on the object plane. It is important to express the total or full field of view rather than a subset of the field of view. This is an extremely critical point to remember. For example, assume we have a CCD camera lens covering a sensor with a $3 \times 4 \times 5$ aspect ratio. We could specify the *horizontal* field of view, which is often

done in video technology and cinematography. However, if we do this, we would be ignoring the full diagonal of the field of view. If you do specify a field of view less than the full or total field, you absolutely must indicate this. For example, it is quite appropriate to specify the field of view as $\pm 10^\circ$. This means, of course, that the total or full diagonal field of view is 20° . Above all, do not simply say “field of view 10°” as the designer will be forced to guess what you really mean!

System specifications should include a defined *spectral range* or wavelength band over which the system will be used. A visible system, for example, generally covers the spectral range from approximately 450 nm to 650 nm. It is important to specify from three to five specific wavelengths and their corresponding relative weights or importance factors for each wavelength. If your sensor has little sensitivity, say, in the blue, then the image quality or performance of the optics can be more degraded in the blue without perceptible performance degradation. In effect, the spectral weights represent an importance factor across the wavelength band where the sensor is responsive. If we have a net spectral sensitivity curve, as in Fig. 1.4, we first select five representative

Figure 1.4
Example of Spectral
Sensitivity Curve



wavelengths distributed over the band, $\lambda_1 = 450$ nm through $\lambda_5 = 650$ nm, as shown. The circular data points represent the relative sensitivity at the specific wavelengths, and the relative weights are now the normalized area or integral within each band from band 1 through band 5, respectively. Note that the weights are not the ordinate of the curve at each wavelength as you might first expect but rather the integral within each band. Table 1.1 shows the data for this example.

Even if your spectral band is narrow, you must work with its bandwidth and derive the relative weightings. You may find some cases where you think the spectral characteristics suggest a monochromatic situation but in reality, there is a finite bandwidth. Pressure-broadened spectral lines emitted by high-pressure arc lamps exhibit this characteristic. Designing such a system monochromatically could produce a disastrous result. In most cases, laser-based systems only need to be designed at the specific laser wavelength.

System *packaging constraints* are important to set at the outset of a design effort, if at all possible. These include length, diameter, weight, distance or clearance from the last surface to the image, location and space for fold mirrors, filters, and/or other components critical to the system operation.

Sets of specifications often neglected until it is too late are the *environmental parameters* such as *thermal soak* conditions (temperature range) that the system will encounter. Also, we may have *radial thermal gradients*, which are changes in temperature from the optical axis outward; *diametral thermal gradients*, which are thermal gradients across the diameter of the system in a nonaxially symmetrical profile; and *axial gradients*, which are thermal gradients from the front to the rear of the system. You may also be provided with a set of *operational* specifications and a set of *storage* specifications with respect to temperature.

TABLE 1.1

	Wavelength, nm	Relative sensitivity	Relative weight
Example of Spectral Sensitivity and Relative Wavelength Weights	450	0.05	0.08
	500	0.2	0.33
	550	1.0	1.0
	600	0.53	0.55
	650	0.09	0.16

System *transmittance*, or throughput, as well as *relative illumination*, or brightness, uniformity over the image format are also often specified.

One of the most important specifications is the optical performance or *image quality*. The following list contains some of the more common ways of specifying the image quality, along with simple definitions. Each of these will be discussed in more detail in Chap. 10.

- *Modulation transfer function (MTF)*. The modulation (think of the word contrast) versus the number of line pairs per millimeter in the image
- *RMS blur diameter*. The diameter of a circle containing approximately 68% of the energy imaged from a point source
- *Encircled energy (or ensquared energy)*. The diameter of a circle (or side of a square such as a pixel) containing a given percent of energy, for example, 80%
- *Root-mean-square (rms) wavefront error*. The rms departure of the real wavefront from a perfect wavefront
- *Other*. Depending on the functional requirements of the system, there may be other performance requirements relating to image quality, for example, point spread function (PSF), control of specific aberrations, etc.

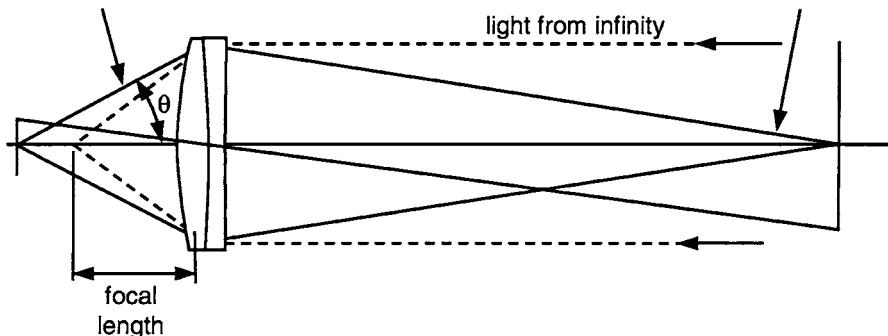
The most fundamental set of first-order specifications are *focal length*, *clear aperture diameter* (more properly called *entrance pupil diameter*, as will be explained in Chap. 2), and *f/number*. As we know, the *f/number* is the focal length divided by the clear aperture diameter (the entrance pupil diameter). There is, however, another important and related quantity called the *numerical aperture* (NA), which is often used. The numerical aperture is simply the sine of the half cone angle of the limiting edge ray coming to the axial image, or the sine of the half cone angle coming from the axial object point, as shown in Fig. 1.5. Why would we want or need yet another term to remember? The reason is that the definition of focal length is based on light from infinity entering the system. What if we have a so-called finite conjugate system where neither the object nor the image is at infinity? The traditional definition of focal length and *f/#* would be misleading since the system really is not being used with collimated light input. Numerical aperture is the answer. The *numerical aperture* is simply the sine of the image cone half angle, regardless of where the object is located. We can also talk about the numerical aperture at the object, which is the sine of the half cone angle from

Figure 1.5

Numerical Aperture
and $f/\#$

the NA is the sine
of the limiting
ray angle θ

light from a
finite object
distance



	<u>f/#</u>	<u>NA</u>
	1	0.5
<i>Focal length implies light from infinity</i>	1.5	0.3333
	2	0.25
	4	0.125
	8	0.0625

the optical axis to the limiting marginal ray emanating from the center of the object. Microscope objectives are routinely specified in terms of numerical aperture. Some microscope objectives reimagine the object at a finite distance, and some have collimated light exiting the objective. These latter objectives are called *infinity corrected* objectives, and they require a “tube lens” to focus the image into the focal plane of the eyepiece or alternatively onto the CCD or other sensor.

As noted earlier, the definition of focal length implies light from infinity. And similarly, f/number is focal length divided by the clear aperture diameter. Thus, f/number is also based on light from infinity. Two terms commonly encountered in finite conjugate systems are “ f/number at used conjugate” and “working f/number .” These terms define the equivalent f/number , even though the object is not at infinity. The f/number at used conjugate is $1/(2 \cdot \text{NA})$, and this is valid whether the object is at infinity or at a finite distance.

It is important at the outset of a design project to compile a specification for the desired system and its performance. The following is a candidate list of specifications:

Optical system basic operational and performance specifications and requirements

Basic system parameters:	
Object distance	
Image distance	
Object to image total track	
Focal length	
f/number (or numerical aperture)	
Entrance pupil diameter	
Wavelength band	
Wavelengths and weights for 3 or 5 λ s	
Full field of view	
Magnification (if finite conjugate)	
Zoom ratio (if zoom system)	
Image surface size and shape	
Detector type	
Optical performance:	
Transmission	
Relative illumination (vignetting)	
Encircled energy	
MTF as a function of line pairs/mm	
Distortion	
Field curvature	
Lens system:	
Number of elements	
Glass versus plastic	
Aspheric surfaces	
Diffractive surfaces	
Coatings	
Sensor:	
Sensor type	
Full diagonal	
Number of pixels (horizontal)	
Number of pixels (vertical)	
Pixel pitch (horizontal)	
Pixel pitch (vertical)	
Nyquist frequency at sensor, line pairs/mm	
Packaging:	
Object to image total track	
Entrance and exit pupil location and size	
Back focal distance	
Maximum diameter	
Maximum length	
Weight	
Environmental:	
Thermal soak range to perform over	
Thermal soak range to survive over	
Vibration	
Shock	
Other (condensation, humidity, sealing, etc.)	

Optical system basic operational and performance specifications and requirements
(Continued)

Illumination:	
Source type	
Power, in watts	
Radiometry issues, source:	
Relative illumination	
Illumination method	
Veiling glare and ghost images	
Radiometry issues, imaging:	
Transmission	
Relative illumination	
Stray light attenuation	
Schedule and cost:	
Number of systems required	
Initial delivery date	
Target cost goal	

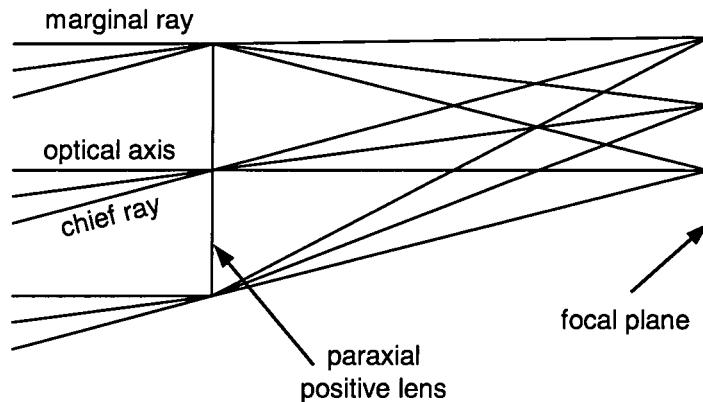
Basic Definition of Terms

There is a term called *first-order optics*. In first-order optics the bending or refraction of a lens or lens group happens at a specific plane rather than at each lens surface. In first-order optics, there are no aberrations of any kind and the imagery is perfect, by definition.

Let us first look at the simple case of a perfect thin positive lens often called a *paraxial lens*. The limiting aperture that blocks the rays beyond the lens clear aperture is called the *aperture stop*. The rays coming from an infinitely distant object that passes through the lens clear aperture focus in the image plane. A paraxial positive lens is shown in Fig. 1.6. The rays coming from an infinitely distant point on the optical axis approach the lens as the bundle parallel to the optical axis. The ray that goes along the optical axis passes through the lens without bending. However, as we move away from the axis, rays are bent more and more as we approach the edge of the clear aperture. The ray that goes through the edge of the aperture parallel to the optical axis is called the *marginal ray*. All of the rays parallel to the optical axis focus at a point on the optical axis in the focal plane. The rays that are coming from a nonaxial object point form an angle with the optical axis. One of these rays is called a *chief ray*, and it goes through the center of the lens (center of the *aperture stop*) without bending.

Figure 1.7 shows a *paraxial lens* (a), and a real lens (b). Both lenses have the same focal length and *f*/number; however the paraxial lens has zero spherical aberration while the real lens has significant spherical aberration.

Figure 1.6
Paraxial Positive Lens



A common first-order representation of an optical system is shown in Fig. 1.8. What we have here is the representation of any optical system, yes, any optical system! It can be a telescope, a microscope, a submarine periscope, or any other imaging optical system.

The easiest way to imagine what we have here is to think of having a shoebox with a 2-in-diameter hole in each end and inside is some arbitrary optical system (or perhaps nothing at all!). If we send a laser beam into the shoebox through the center of the left-hand hole normal to the hole, it will

Figure 1.7
Paraxial Lens (a) and
Real Lens (b)

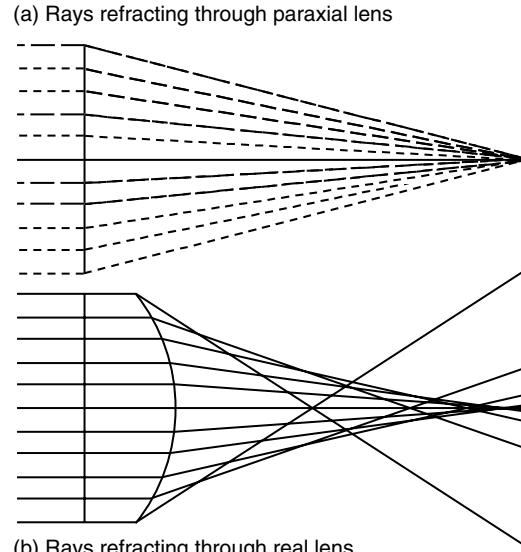
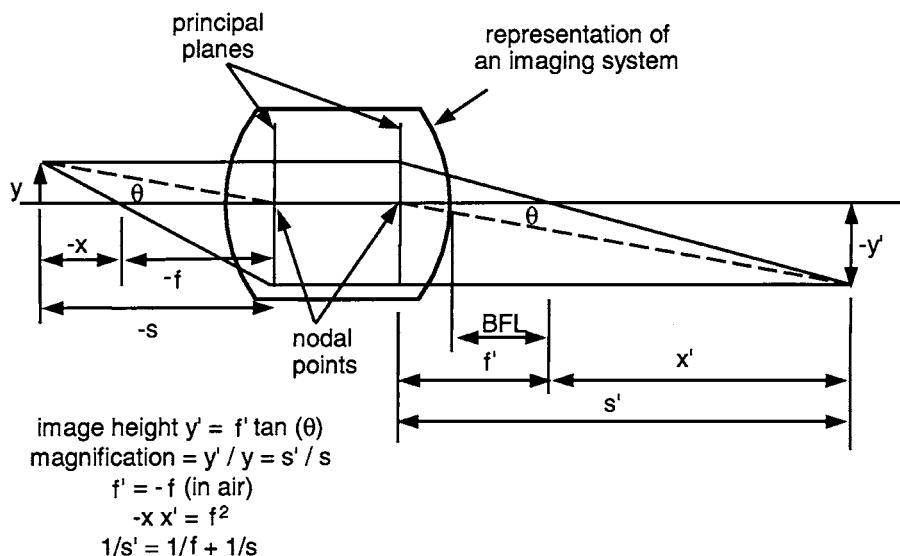


Figure 1.8
Cardinal Points of an Optical System



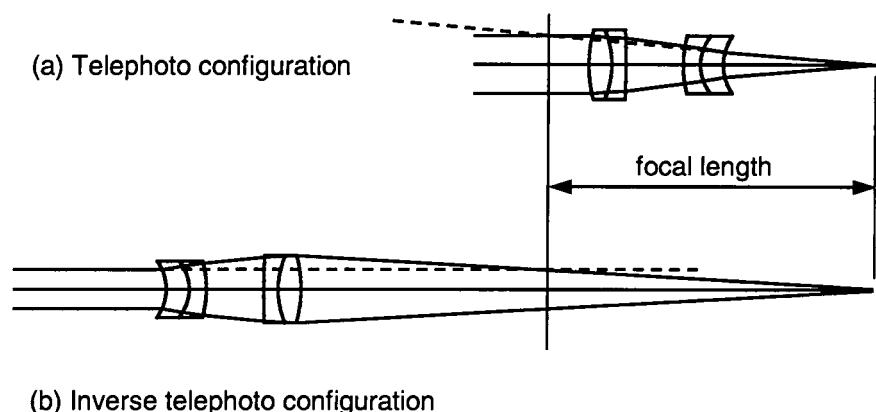
likely exit through the center of the hole at the other end of the shoebox. The line going through the center of each of the holes is the optical axis.

If we now send the laser beam into the shoebox displaced nearly 1 in vertically, it may exit the shoebox on the other end exactly the same and parallel to how it entered, in which case there is probably nothing in the shoebox. Alternately, the laser beam may exit the shoebox either descending or ascending (going downhill or uphill). If the laser beam is descending, it will cross the optical axis somewhere to the right of the shoebox, as shown in Fig. 1.8. If we connect the entering laser beam with the exiting laser beam, they will intersect at a location called the *second principal plane*. This is sometimes called the *equivalent refracting surface* because this is the location where all of the rays appear to bend about. In a high-performance lens, this equivalent refracting surface is spherical in shape and is centered at the image. The distance from the second principal plane to the plane where the ray intersects the optical axis is the *focal length*.

If we now send a laser beam into the hole on the right parallel to the optical axis and in a direction from right to left, it will exit either ascending or descending (as previously), and we can once again locate the principal plane, this time the *first principal plane*, and determine the focal length. Interestingly, the focal length of a lens system used in air is identical whether light enters from the left or the right. Figure 1.9a shows a telephoto lens whose focal length is labeled. Recall that we can compute

Figure 1.9

The Identical Lens
Showing How the
Focal Length Is Iden-
tical When the Lens Is
Reversed



the focal length by extending the marginal ray back from the image until it intersects the incoming ray, and this distance is the focal length. In the telephoto lens the focal length is longer than the physical length of the lens, as shown. Now consider Fig. 1.9b, where we have taken the telephoto lens and simply reversed it with no changes to radii or other lens parameters. Once again, the intersection of the incoming marginal ray with the ray extending forward from the image is the focal length. The construction in Fig. 1.9b shows clearly that the focal lengths are identical with the lens in either orientation!

The center of the principal planes (where the principal planes cross the optical axis) are called the *nodal points*, and for a system used in air, these points lie on the principal planes. These nodal points have the unique property that light directed at the front nodal point will exit the lens from the second nodal point at exactly the same angle with respect to the optical axis. This, too, we can demonstrate with our laser beam and shoebox.

So far, we have not talked about an object or an image at all. We can describe or represent a cone of light leaving an object (at the height, y , in Fig. 1.8.) as including the ray parallel to the optical axis, the ray aimed at the front nodal point, and lastly the ray leaving the object and passing through the focal point on the left side of the lens. All three of these rays (or laser beams) will come together once again to the right of the lens a distance, y' , from the optical axis, as shown. We will not bore you with the derivation, but rest assured that it does happen this way.

What is interesting about this little example is that our shoebox could contain virtually any kind of optical system, and all of the preceding will hold true. In the case where the laser beam entering parallel to the

optical axis exits perhaps at a different distance from the axis but parallel to the axis, we then have what is called an *afocal lens* such as a laser beam expander, an astronomical telescope, or perhaps a binocular. An afocal lens has an infinite focal length, meaning that both the object and the image are at infinity.

Useful First-Order Relationships

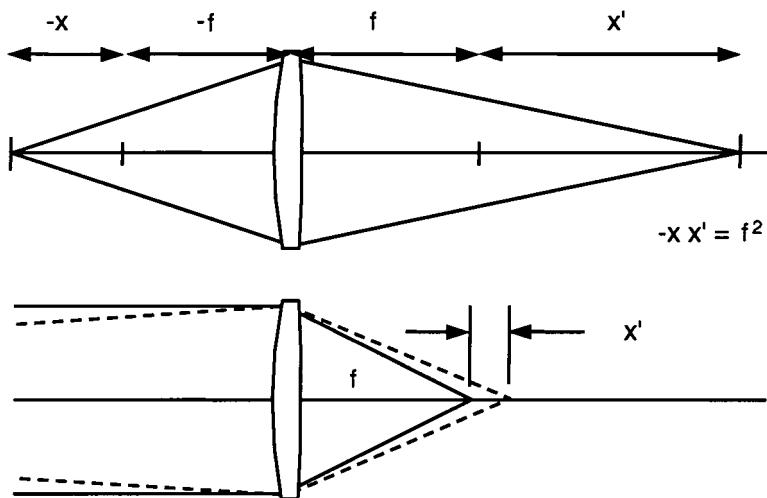
As discussed earlier, in first-order optics, lenses can be represented by planes where all of the bending or refraction takes place. Aberrations are nonexistent in first-order optics, and the imagery is by definition absolutely perfect. There are a series of first-order relationships or equations, which come in very handy in one's everyday work, and we will discuss the most useful ones here.

Consider the simple lens system shown in Fig. 1.10. Newton's equation says:

$$(-x)(x') = f^2$$

where x is the distance from the focal point on the front side of the lens to the object, and x' is the distance from the rear focal point to the image. Note that x is negative according to the sign convention, since the distance from the image to the object is in a direction to the

Figure 1.10
Newton's Equation



left. This is an interesting equation in that, at first glance, it seems to be of marginal use. However, consider the example where we need to determine how far to refocus a 50-mm focal length lens for an object at a distance of 25 m. The result is 0.1 mm, and this is, in all likelihood, a very reliable and accurate answer. We must always remember, however, that first-order optics is an approximation and assumes no aberrations whatsoever. For small angles and large $f/\#$ s the results are generally reliable; however, as the angles of incidence on surfaces increase, the results become less reliable. Consider Fig. 1.11a where we show how light proceeds through a three-element lens known as a Cooke triplet, with the object at infinity. If we were to use Newton's equation to determine how far to refocus the lens for a relatively close object distance, as shown in Fig. 1.11b, the resulting amount of refocusing may not be reliable. This is because the ray heights and angles of incidence are different from the infinite object condition, especially at the outer positive elements, as shown in Fig. 1.11c, which is an overlay of the infinite and close object distance layouts. These different ray heights and angles of

Figure 1.11
Light Imaging
through a Cooke
Triplet for Two Object
Distances

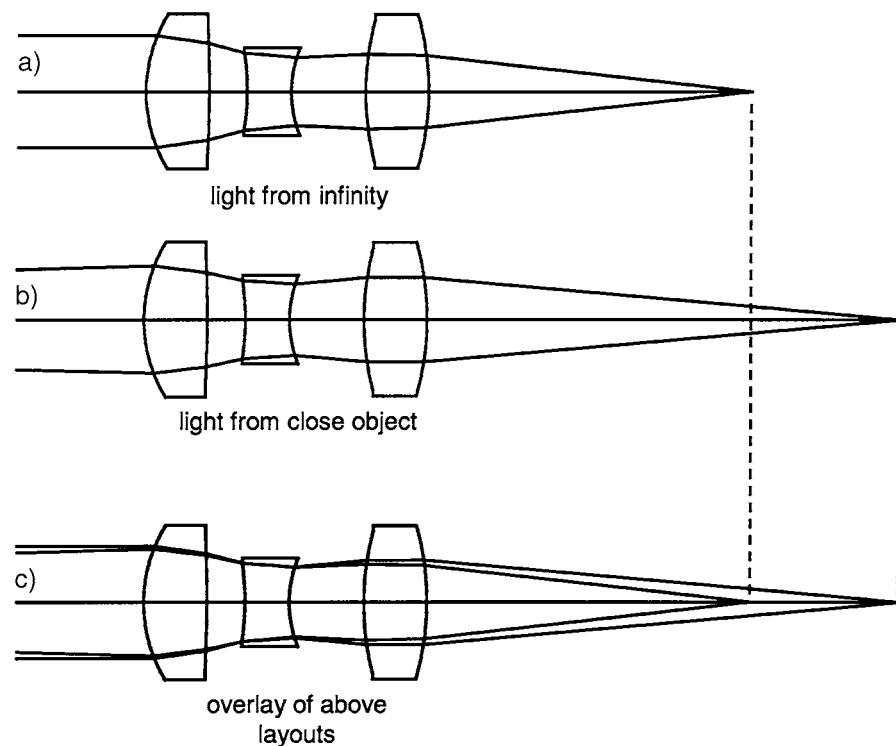
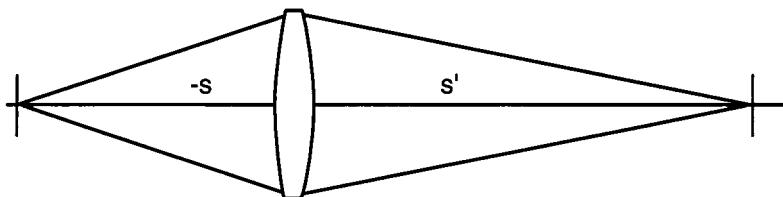


Figure 1.12

Basic Relationship of Object and Image:
The "Lens Makers" Equation



incidence will cause aberrations, and the net effect is that the result determined by Newton's equation might not be reliable for predicting where best focus is located. Consider a typical $f/5$ 50-mm focal length Cooke triplet lens used at an object distance of 0.5 m. Newton's equation gives a required refocusing of 2.59 mm from infinity focus, versus 3.02 mm based on optimum image quality, a difference of 0.43 mm. However, for a 10-m object distance, the difference between Newton's equation and best focus reduces to 0.0008 mm, which is negligible.

The important message here is to use first-order optics with caution. If you have any question as to its accuracy in your situation, you really should perform a computer analysis of what you are modeling. If you then find that your first-order analysis is sufficiently accurate, continue to use it with confidence in similar situations. However, if you find inaccuracies, you may need to work with real rays in your computer model.

Another useful and commonly used equation is

$$\frac{1}{s'} = \frac{1}{f} + \frac{1}{s}$$

where s and s' are the object and image distances, respectively, as shown in Fig. 1.12.

Consider now the basic definitions of magnification from an object to an image. In Fig. 1.13, we show how *lateral magnification* is defined. Lateral implies in the plane of the object or the image, and lateral magnification

Figure 1.13

Lateral Magnification

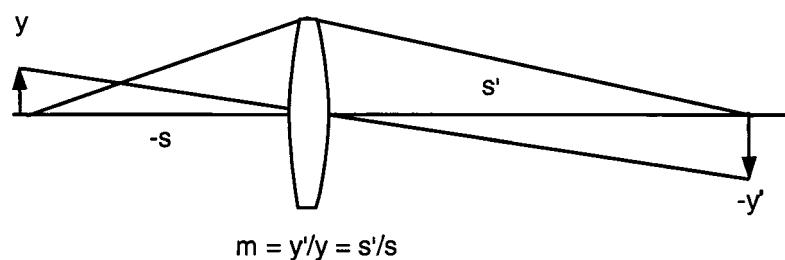
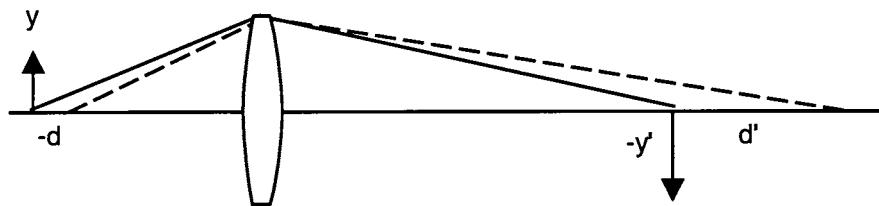


Figure 1.14
Longitudinal Magnification



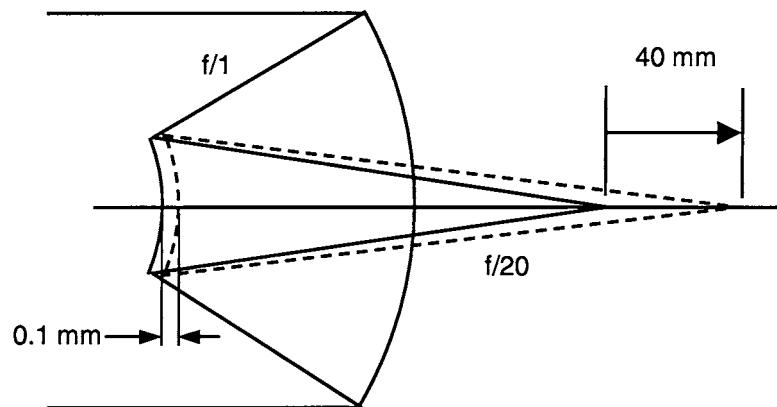
$$\text{longitudinal magnification} = d'/d = (\text{lateral magnification})^2 = (y'/y)^2$$

is therefore the image height, y' divided by the object height, y . It is also the image distance, d' , divided by the object distance, d .

There is another form of magnification: the *longitudinal magnification*. This is the magnification *along the optical axis*. This may be a difficult concept to visualize because the image is always in a given plane. Think of longitudinal magnification this way: if we move the object a distance, d , we need to move the image, d' , where d'/d is the longitudinal magnification. It can be shown that the longitudinal magnification is the square of the lateral magnification, as shown in Fig. 1.14. Thus, if the lateral magnification is $10\times$, the longitudinal magnification is $100\times$. A good example is in the use of an overhead projector where the viewgraph is in the order of 250 mm wide and the screen is in the order of 1 m wide, giving a lateral magnification of $4\times$. If we were to move the viewgraph 25 mm toward the lens, we would need to move the screen outward by $16 \times 25 = 400$ mm.

As a further example of the concept, consider Fig. 1.15 where we show a two-mirror reflective system called a *Cassegrain*. Let us assume

Figure 1.15
Cassegrain Reflective System

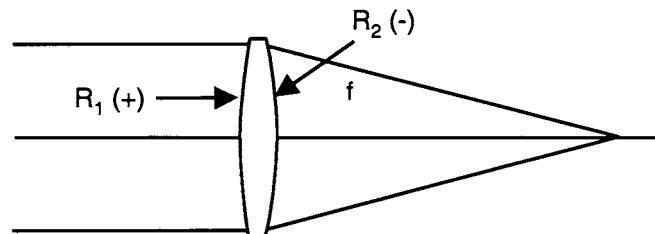


that the large or *primary mirror* is 250 mm in diameter and is $f/1$. Also, assume that the final image is $f/20$. The small, or *secondary*, mirror is, in effect, magnifying the image, which would be formed by the primary mirror by $20\times$ in lateral magnification. Thus, the longitudinal magnification is $400\times$, which is the square of the lateral magnification. Now let us move the secondary mirror 0.1 mm toward the primary mirror. How far does the final image move? The answer is $0.1 \times 400 = 40$ mm to the right. This is a very large amount and it illustrates just how potent the longitudinal magnification really can be.

While we are on the subject, how can we easily determine which way the image moves if we move the secondary mirror to the right as discussed previously? Indeed there is an easy way to answer this question (and similar questions). The approach to follow when presented by a question of this kind is to consider moving the component a very large amount, perhaps even to its limit, and ask "what happens?" For example, if we move the secondary mirror to a position approaching the image formed by the primary, clearly the final image will coincide with the secondary mirror surface when it reaches the image formed by the primary. This means that the final image will move in the same direction as the secondary mirror motion. In addition, if you take the secondary and move it a large amount toward the primary, eventually the light will reflect back to the primary when the rays are incident normal to the secondary mirror surface. Moreover, at some intermediate position, the light will exit to the right collimated or parallel. The secret here, and for many other similar questions, is to *take the change to the limit*. Take it to a large enough magnitude so that the direction of the result becomes fully clear and unambiguous.

Figure 1.16 shows how the *optical power* of a single lens element is defined. The optical power is given by the Greek letter, Φ , and Φ is the

Figure 1.16
Optical Power and
Focal Length of a
Single Lens Element



$$\Phi = \text{Optical power} = \frac{1}{f} = (n - 1) \cdot \left(\frac{1}{r_1} - \frac{1}{r_2} \right)$$

reciprocal focal length or 1 divided by the focal length. In optics, we use a lot of reciprocal relationships. Power = $\Phi = 1/\text{focal length}$, and curvature = $(1/\text{radius})$ is another.

If we know the radii of the two surfaces, r_1 and r_2 , and the refractive index, n , we find that

$$\Phi = \frac{1}{\text{focal length}} = (n - 1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right)$$

In addition, if we have two thin lenses separated by an air space of thickness d , we find that

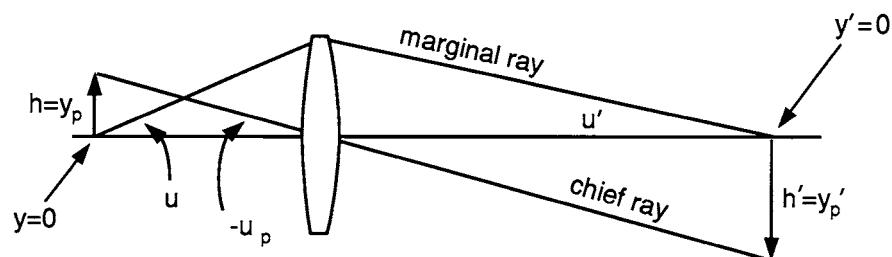
$$\Phi = \frac{1}{\text{focal length}} = \Phi_a + \Phi_b - d(\Phi_a \Phi_b)$$

One very important constant in the optical system is the *optical invariant* or *Lagrange invariant* or *Helmholtz invariant*. It has a constant value throughout the entire system, on all surfaces and in the spaces between them. The optical invariant defines the system throughput. The basic characteristic of an optical system is known when the two main rays are traced through the system: the *marginal ray* going from the center of the object through the edge of the aperture stop and the *chief* or *principal ray* going from the edge of the object through the center of the aperture stop. These rays are shown in Fig. 1.17. The optical invariant defines the relationship between the angles and heights of these two rays through the system, and in any location in the optical system it is given as

$$I = y_p n u - y n u_p$$

where the subscript p refers to the principal ray, no subscript refers to the marginal ray, and n is the refractive index.

Figure 1.17
The Optical Invariant



The optical invariant, I , once computed for a given system, remains constant everywhere within the system. When this formula is used to calculate the optical invariant in the object plane and in the image plane where the marginal ray height is zero, then we get the commonly used form of the optical invariant

$$I = hnu = h'n'u'$$

where h , n , and u are the height of the object, the index of refraction, and angle of the marginal ray in the object plane, and h' , n' , and u' are the corresponding values in the image space. Although this relationship is strictly valid only in the paraxial approximation, it is often used with sufficient accuracy in the form

$$nh \sin u = n'h' \sin u'$$

From this form of optical invariant we can derive the magnification of the system $M = h'/h$ as

$$M = \frac{n \sin u}{n' \sin u'}$$

In simple terms these relationships tell us that if the optical system magnifies or increases the object M times, the viewing angle will be decreased M times.

In systems analysis, the specification of the optical invariant has a significant importance. In the radiometry and photometry of an optical system, the total light flux collected from a uniformly radiating object is proportional to I^2 of the system, commonly known as *etendue*, where I is the optical invariant. For example, if the optical system is some kind of a projection system that uses a light source, then the projection system with its optical invariant defines the light throughput. It is useful to compare the optical invariant of the light source with the invariant of the system to see how much light can be coupled into the system. It is not necessarily true that the choice of a higher-power light source results in a brighter image. It can happen that the light-source optical invariant is significantly larger than the system optical invariant, and a lot of light is stopped by the system. The implications of the optical invariant and etendue on radiometry and photometry will be discussed in more depth in Chap. 14.

The *magnification* of a visual optical system is generally defined as the ratio of the angles subtended by the object *with* or looking through

the optical system to the angle subtended by the object *without* the optical system or looking at the object directly with unaided vision. In visual optical systems where the human eye is the detector, a nominal viewing distance without the optical system when the magnification is defined as unity is 250 mm. The reason that unity magnification is defined at a distance of 250 mm is that this is the closest distance that most people with good vision can focus. As you get closer to the object, it subtends a larger angle and hence looks bigger or magnified.

This general definition of magnification takes different forms for different types of optical systems. Let us look first at the case of a microscope objective with a CCD camera, as shown in Fig. 1.18. The image from the CCD is brought to the monitor and the observer is located at the distance, D , from the monitor. The question is what is the magnification of this system. In the first step a microscope objective images the object with the height, y , onto the CCD camera, with the magnification

$$\frac{y'}{y}$$

where y' is the image height at the CCD camera. In the next step the image from the CCD is brought to the monitor with the magnification

$$\frac{y''}{y'}$$

Figure 1.18
Magnification of
a Microscope

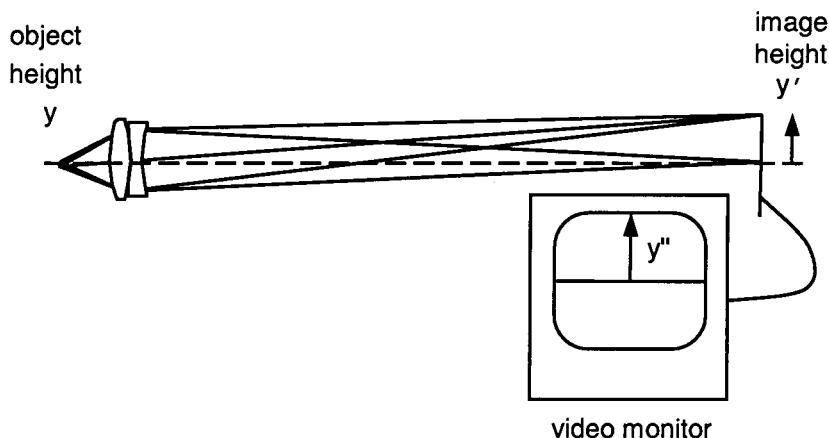
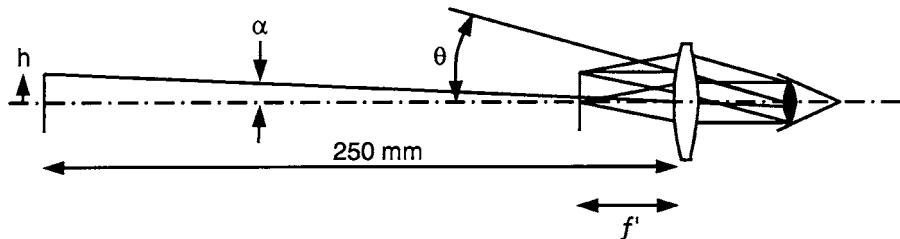


Figure 1.19
Magnification of a Magnifier or Eyepiece



where y'' is the image height at the monitor. In the third step the observer watches the monitor from the distance, D , with the magnification

$$\frac{250 \text{ mm}}{D}$$

Overall, the magnification of this system is

$$M = \frac{y'}{y} \frac{y''}{y'} \frac{250}{D}$$

$$M = \frac{y''}{y} \frac{250}{D}$$

The second example is a magnifier or an eyepiece, as shown in Fig. 1.19. The object with height h at a distance of 250 mm is seen to subtend an angle, α . When the same object is located in the first focal plane of the eyepiece, the eye sees the same object at an angle, θ , where

$$\alpha = \frac{h}{250} \quad \theta = \frac{h}{f}$$

Therefore, magnification M is given by

$$M = \frac{\theta}{\alpha} = \frac{250}{f}$$

The next example is the visual microscope shown in Fig. 1.20. A microscope objective is a short focal length lens, which forms a highly

Figure 1.20
Magnification of a Visual Microscope

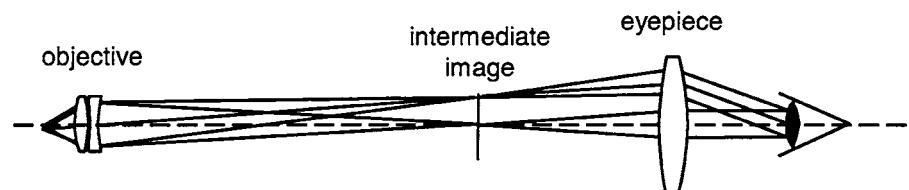
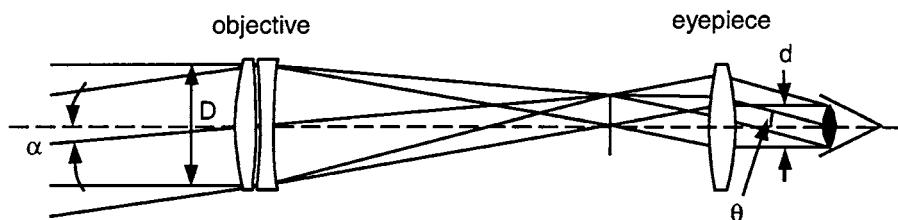


Figure 1.21

Magnification of a Visual Telescope



magnified image of the object. A visual microscope includes an eyepiece which has its front focal plane coincident with the objective image plane. The image formed by the objective is seen through the eyepiece, which has its magnification defined as

$$M_e = \frac{250}{f}$$

where f is the focal length of the eyepiece. The magnification of the microscope is the product of the magnification of the objective times the magnification of the eyepiece. Thus

$$M_m = M_o M_e$$

A visual telescope is shown in Fig. 1.21. A distant object is seen at an angle, α , without the telescope and at an angle, θ , with the telescope. The angular magnification of the telescope is

$$M = \frac{\theta}{\alpha}$$

Using the similarity of triangles, it can also be shown that the telescope magnification is

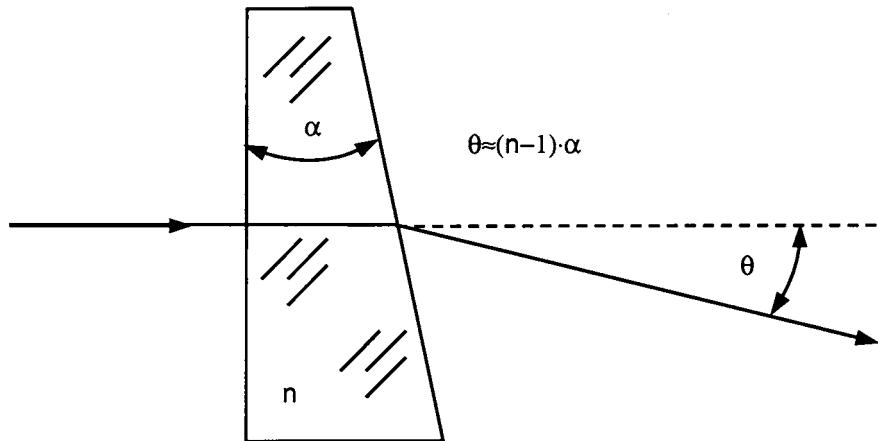
$$M = \frac{f_o}{f_e} = \frac{D}{d}$$

where f_o is the focal length of the objective, f_e is the focal length of the eyepiece, D is the diameter of the entrance pupil, and d is the diameter of the exit pupil.

There are several useful first-order relationships regarding plane parallel plates in an optical system. The first relates to what happens in an optical system when a wedge is added to a plane parallel plate. If a ray, as shown in Fig. 1.22, goes through the wedged piece of material of index

Figure 1.22

Light Deviation through Wedged Material



of refraction n and a small wedge angle α , the ray deviates from its direction of incidence by the angle, θ , according to

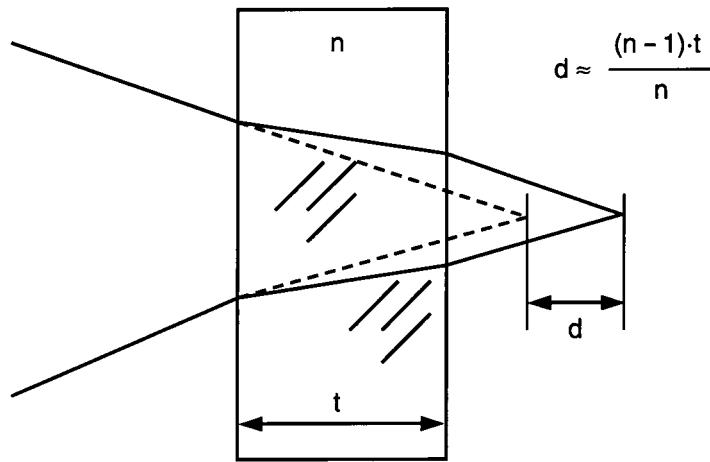
$$\theta \approx (n - 1)\alpha$$

The angle of deviation depends on the wavelength of light, since the index of refraction is dependent on the wavelength. It is important to understand how the wedge can affect the performance of the optical system. When a parallel beam of white light goes through the wedge, the light is dispersed into a rainbow of colors, but the rays of the individual wavelengths remain parallel. Therefore, the formula that gives the angle of deviation through the wedge is used to quickly determine the allowable wedge in protective windows in front of the optical system. However, if the wedge is placed into a converging beam, not only will the different colors be focused at different distances from the optical axis, but also the individual colors will be blurred. There is a term called *boresight error*, which means the difference between where you think the optical system is looking and where it really is looking. A wedged window with a wedge angle, α , will cause a system to have a boresight error of angle θ .

A plane parallel plate in a converging beam moves the image plane further along the optical axis, as shown in Fig. 1.23. If the thickness of the plate is t , the image displacement, d , along the optical axis is

$$d = (n - 1) \frac{t}{n}$$

Figure 1.23
Focus Shift
Introduced by a
Plane Parallel Plate

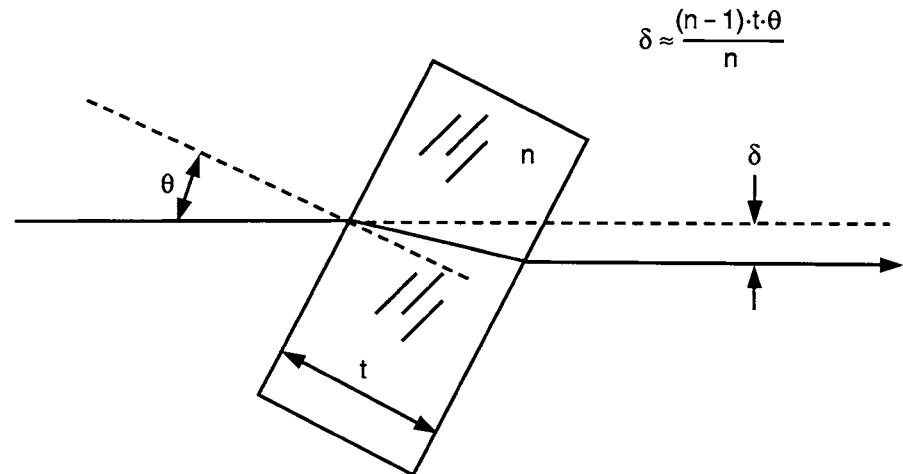


When a plane parallel plate is tilted in the optical system, as in Fig. 1.24, then the ray incident at an angle, θ , is displaced laterally by the amount, δ , given by

$$\delta = (n - 1) \frac{t\theta}{n}$$

Note that if we look through a telescope at an infinitely distant object and we put a tilted plane parallel plate in front of the telescope,

Figure 1.24
Lateral Displacement
of a Ray Introduced
by a Tilted Plane
Parallel Plate



there will be no change in the image. One would think that color fringes would be seen because the different wavelengths are displaced differently. However, because the parallel bundle of rays going through the tilted plate is only laterally displaced, it remains parallel to itself after transmission through the plate, and therefore there is no color fringing. If there is a wedge in the plate, however, chromatic dispersion will, of course, cause the appearance of color fringing.

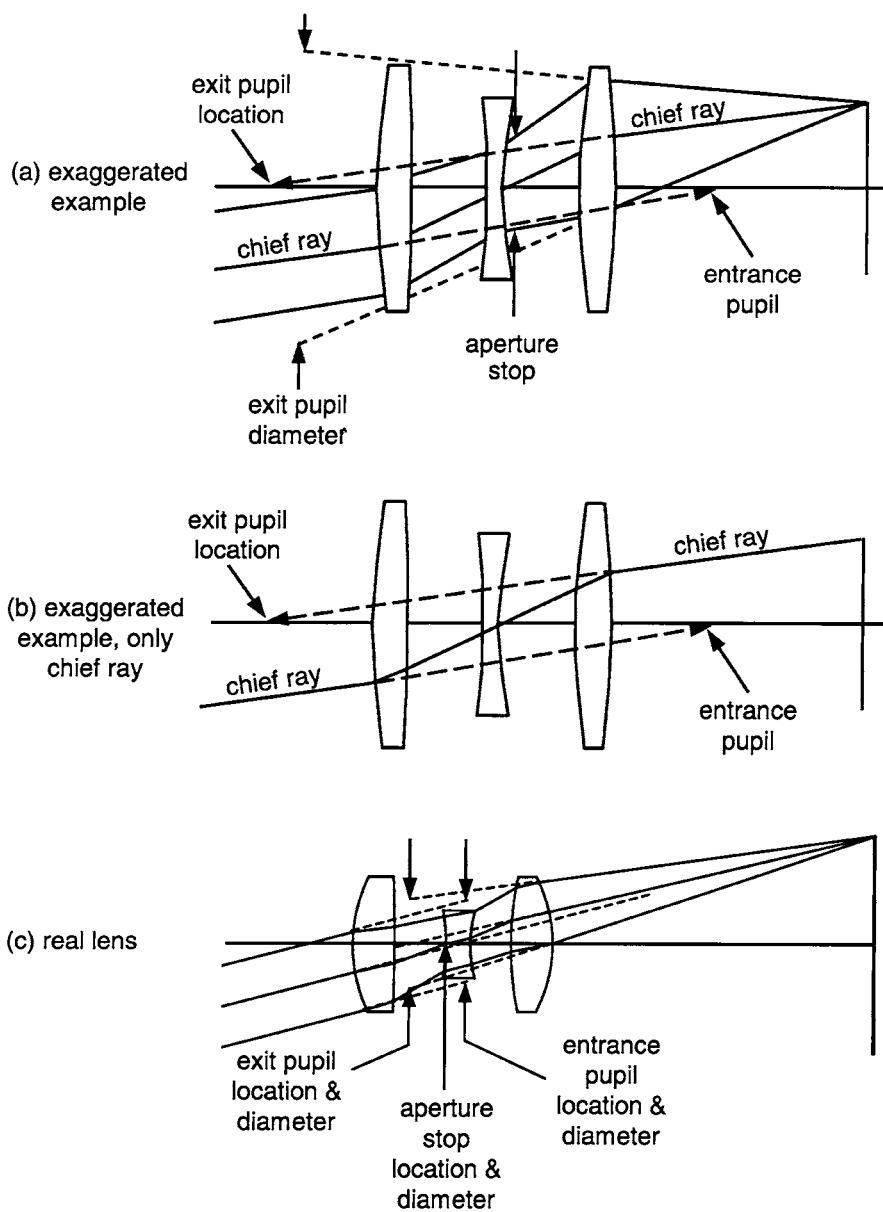
Stops and Pupils and Other Basic Principles

The Role of the Aperture Stop

In an optical system, there are apertures which are usually circular that limit the bundles of rays which go through the optical system. In Fig. 2.1 a classical three-element form of lens known as a Cooke triplet is shown as an example. Take the time to compare the exaggerated layout (Figs. 2.1a and b) with an actual computer optimized design (Fig. 2.1c). From each point in the object only a given group or bundle of rays will go through the optical system. The *chief ray*, or *principal ray*, is the central ray in this bundle of rays. The *aperture stop* is the surface in the system where all of the chief rays from different points in the object cross the optical axis and appear to pivot about. There is usually an iris or a fixed mechanical diaphragm or aperture in the lens at this location. If your lens has an adjustable iris at the stop, its primary purpose is to change the brightness of the image. The chief ray is, for the most part, a mathematical convenience; however, there definitely is a degree of symmetry that makes its use valuable. We generally refer to the specific height of the chief ray on the image as the image height.

Figure 2.1

Aperture Stop and Pupils in an Optical System



Entrance and Exit Pupils

The *entrance pupil* is the image of the aperture stop when viewed from the front of the lens, as shown in Fig. 2.1. Indeed, if you take any telescope, such as a binocular, and illuminate it from the back and look into the optics from the front, you will see a bright disk which is formed, in most cases, at the objective lens at the front of the binocular. In the opposite case, if you illuminate the system from the front, there will be a bright disk formed behind the eyepiece. The image of the aperture stop in the image space is called the *exit pupil*. If you were to write your initial with a grease pencil on the front of the objective lens and locate a business card at the exit pupil, you would see a clear image of the initial on the card.

There is another way to describe entrance and exit pupils. If the chief ray entering the lens is extended without bending or refracting by the lens elements, it will cross the optical axis at the entrance pupil. This is shown in Figs. 21a and b where only the chief ray and the pupil locations are shown for clarity. Clearly, it is the image of the aperture stop since the chief ray crosses the optical axis at the aperture stop. In a similar way, the exit pupil will be at the location where the chief ray appears to have crossed the optical axis. The location of the exit pupil can be obtained if the chief ray that exits the optical system is extended backwards until it crosses the optical axis. Both definitions are synonymous, and it will be valuable to become familiar with each.

Let us assume that we have an optical system with a lot of optical components or elements, each of them having a known clear aperture diameter. There are also a few mechanical diaphragms in the system. The question is, which of all these apertures is the aperture stop? In order to answer this question, we have to image each aperture into object space. The aperture whose image is seen from the object plane at the smallest angle is the aperture stop. It is the limiting aperture within the lens.

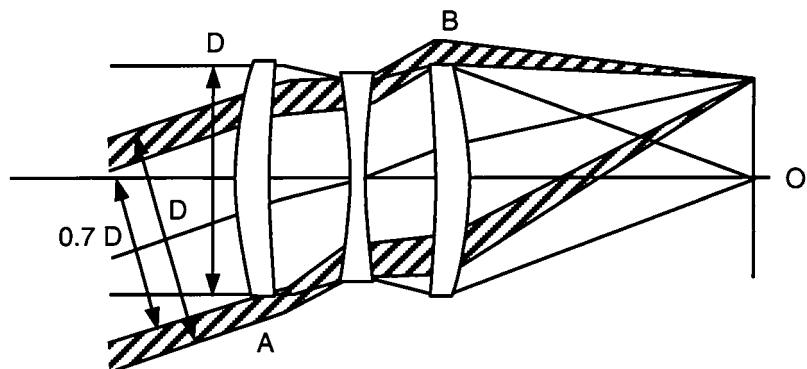
There are many systems such as stand-alone camera lenses, where the location of the entrance and exit pupils are generally not important. The exit pupil location of a camera lens will, of course, dictate the angle of incidence of the off-axis light onto the sensor. However, the specific pupil locations are generally not functionally critical. When multiple groups of lenses are used together, then the pupil locations become very important since the exit pupil of one group must match the entrance pupil location of the following group. This will be discussed later in this chapter.

Vignetting

The position of the aperture stop and the entrance and exit pupils is very important in optical systems. Two main reasons will be mentioned here. The first reason is that the correction of aberrations and image quality very much depends on the position of the pupils. This will be discussed in detail later in the book. The second reason is that the amount of light or throughput through the optical system is defined by the pupils and the size of all elements in the optical system. If ray bundles from all points in the field of view fill the aperture stop entirely and are not truncated or clipped by apertures fore or aft of the stop, then there is no *vignetting* in the system.

For a typical lens, light enters the lens on axis (the center of the field of view) through an aperture of diameter D in Fig. 2.2, and focuses down to the center of the field of view. As we go off axis to the maximum field of view, we are now entering the lens at an angle. In order to allow the rays from the entire diameter, D , to proceed through the lens, in which case the aperture stop will be filled with the ray bundle from the edge of the field, the rays at the edge of the pupil have to go through points A and B . At these positions, A and B , the rays undergo severe bending which means that they contribute significantly to the image aberrations of the system, as will be discussed in Chaps. 3 and 5. At the same time, mounting of the lenses with larger diameters is more expensive. Further, the lens will be heavier and thicker. So why don't we truncate the aperture in the plane of Fig. 2.2 to $0.7D$? We will lose approximately 30% of the energy at the edge of the field of view compared to the center of the field; however, the positive elements in our Cooke triplet example will be smaller in diameter, which means that they can also be thinner and the housing can be smaller.

Figure 2.2
Vignetting



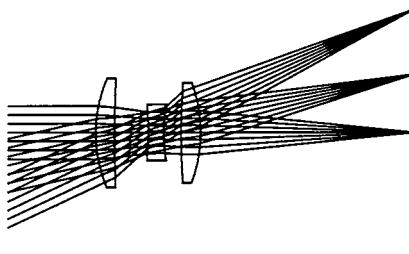
and lighter in weight. Telescopes, projectors, and other visual optical systems can have vignetting of about 30 to 40%, and the eye can generally “tolerate” this amount of vignetting. When we say that the eye can tolerate 30 to 40% vignetting, what we mean is that a slowly varying brightness over an image of this magnitude is generally not noticed. A good example is in overhead viewgraph and slide projectors where this amount of brightness falloff is common, yet unless it is pointed out, most people simply will not notice. If the film in a 35-mm camera has a large dynamic range, then this magnitude of vignetting is also acceptable in film-based photography. In digital cameras the vignetting can be calibrated out since the vignetting is a known function.

In Fig. 2.3 a triplet lens example is shown first in its original form without vignetting (Fig. 2.3a). In the next step the elements are sized for 40% vignetting, but with the rays traced as if there is no vignetting (Fig. 2.3b). In the last step the lens is shown with the vignetted bundle of rays at the edge of the field (Fig. 2.3c).

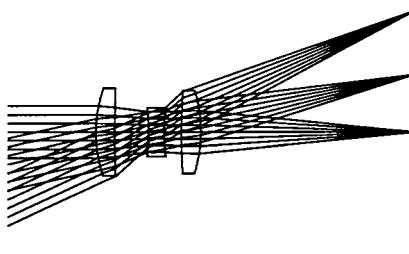
Figure 2.3

Example of
Vignetting

(a) Basic design $f/5 \pm 20$
degree field, no
vignetting



(b) Elements sized for
40% vignetting



(c) Final design with 40%
vignetting

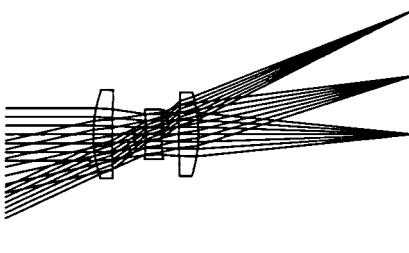
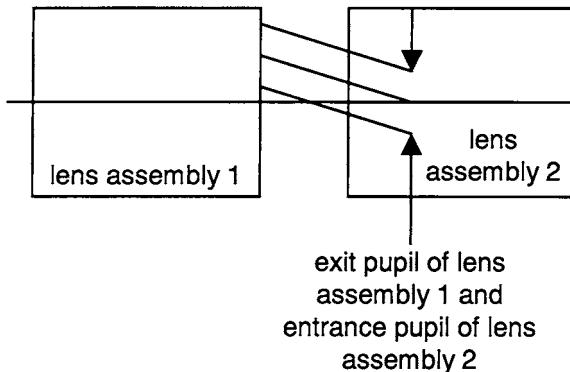


Figure 2.4

Matching of Pupils



Although vignetting is acceptable and often desirable in visible optical systems, it can be devastating in thermal infrared optical systems because of image anomalies, as will be discussed in Chap. 12. One must also be very careful when specifying vignetting in laser systems, as will be discussed in Chap. 11.

When a system is designed using off-the-shelf components with a combination of two or more modules or lens assemblies, it is very important to know the positions of the entrance and exit pupils of these modules. The exit pupil of the first module *must* coincide with the entrance pupil of the second module, etc. This is shown in Fig. 2.4.

There can be a very serious pupil-matching problem when using off-the-shelf (or even custom) zoom lenses as modules in optical systems. Zoom lenses have a given size and position of their pupils which change as a function of zoom position or focal length. It is very easy to make a mistake when the exit pupil of the first module is matched to the entrance pupil of the second module for only one zoom position. When the pupils move with respect to one another through zoom and do not image from one to another, we can lose the entire image.

Diffraction, Aberrations, and Image Quality

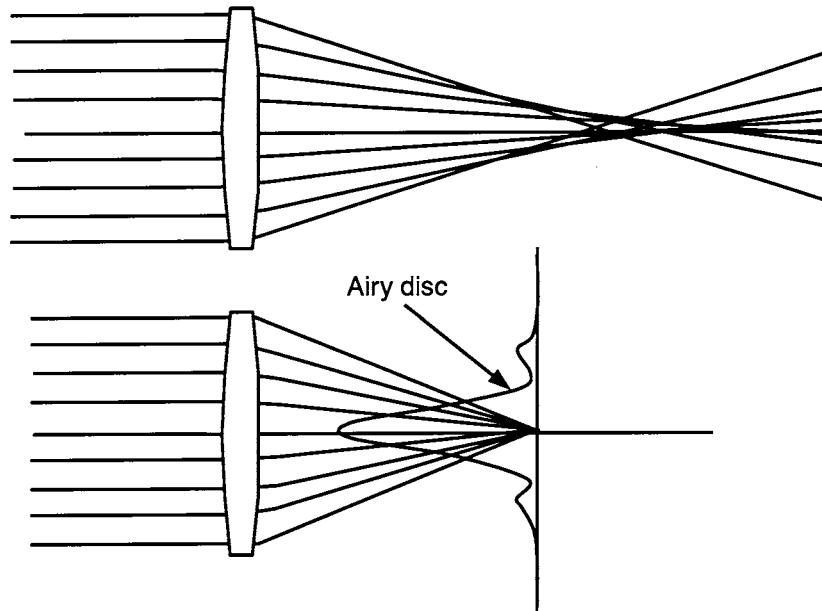
What Image Quality Is All About

Image quality is never perfect! While it would be very nice if the image of a point object could be formed as a perfect point image, in reality we find that image quality is degraded by either geometrical aberrations and/or diffraction. Figure 3.1 illustrates the situation. The top part of the figure shows a hypothetical lens where you can see that all of the rays do not come to a common focus along the optical axis. Rather, the rays entering the lens at its outer periphery cross the optical axis progressively closer to the lens than those rays entering the lens closer to the optical axis. This is one of the most common and fundamental aberrations, and it is known as *spherical aberration*. Geometrical aberrations are due to the failure of the lens or optical system to form a perfect geometrical image. These aberrations are fully predictable to many decimal places using standard well-known ray trace equations.

If there were no geometrical aberrations of any kind, the image of a point source from infinity is called an *Airy disk*. The profile of the Airy

Figure 3.1

Image Quality, Geometrical Aberrations
(Top) and Diffraction Limited (Bottom)



disk looks like a small gaussian intensity function surrounded by low-intensity rings of energy, as shown in Fig. 3.1, exaggerated.

If we have a lens system in which the geometrical aberrations are significantly larger than the theoretical diffraction pattern or blur, then we will see an image dominated by the effect of these geometrical aberrations. If, on the other hand, the geometrical aberrations are much smaller than the diffraction pattern or blur, then we will see an image dominated by the effect of the Airy disk. If we have a situation where the blur diameter from the geometrical aberration is approximately the same size as the theoretical diffraction blur, we will see a somewhat degraded diffraction pattern or Airy disk. Figure 3.1, while exaggerated, does show a situation where the resulting image would, in fact, be a somewhat degraded Airy disk.

What Are Geometrical Aberrations and Where Do They Come From?

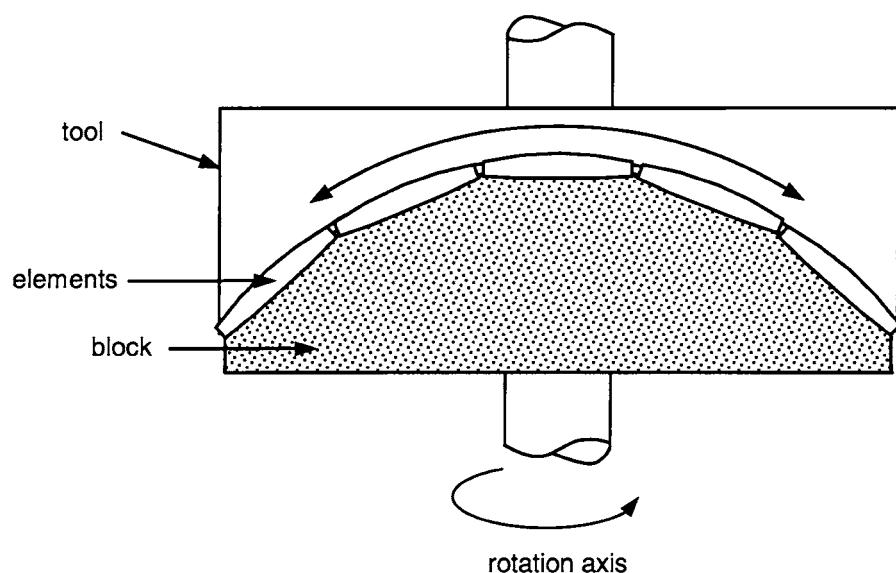
In the previous section, we have shown the distinction between geometrical aberrations and diffraction. The bottom line is that imagery formed by lenses with spherical surfaces simply is not perfect! We use

spherical surfaces primarily because of their ease of manufacture. In Fig. 3.2, we show how a large number of elements can be ground and polished using a common or single tool. The elements are typically mounted to what is called a *block*. Clearly, the smaller the elements and the shallower the radius, the more elements can be mounted on a given block. The upper tool is typically a spherical steel tool. The grinding and polishing operation consists of a rotation about the vertical axis of the blocked elements along with a swinging motion of the tool from left to right, as indicated by the arrows. The nature of a sphere is *that the rate of change of slope is constant everywhere on a sphere*, and because of this mathematical definition, the tool and lens surfaces will only be in perfect contact with one another over the full range of motions involved when both are perfectly spherical. Due to asymmetries in the process, the entire surface areas of the elements and tool are not in contact the same period of time. Hence this process is not perfect. However, the lens surfaces are driven to a near-spherical shape in reasonable time by a skilled optician. This is the reason we use spherical surfaces for most lenses. Chapter 18 discusses optical component manufacturing in more detail.

We will discuss the use of nonspherical or aspheric surfaces in Chap. 8.

Figure 3.2

Manufacture of Spherical Lens Surfaces



Earlier we said that geometrical aberrations are due entirely to the failure of the lens or optical system to form a perfect geometrical image. Maxwell formulated three conditions that have to be met for the lens to form a perfect geometrical image:

1. All the rays from object point O after passing through the lens, must go through the image point O' .
2. Every segment of the object plane normal to the optical axis that contains the point O must be imaged as a segment of a plane normal to the optical axis, which contains O' .
3. The image height, h' , must be a constant multiple of the object height, h , no matter where O is located in the object plane.

Violation of the first condition results in the image degradation, or image aberrations. Violation of the second condition results in the presence of image curvature, and violation of the third condition in image distortion. A different way to express the first condition is that all the rays from the object point, O , must have the same *optical path length* (OPL) to the image point, O' .

$$\text{OPL} = \int_{O(x,y)}^{O'(x',y')} n(s)ds$$

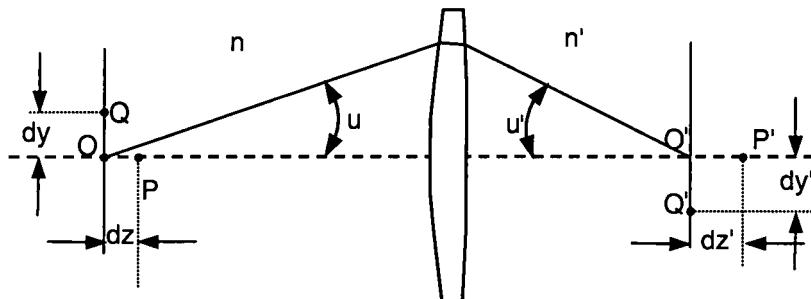
where $n(s)$ is the index of refraction at each point along the ray path, s .

The lenses that meet the first Maxwell condition are called stigmatic. Perfect stigmatic lenses are generally stigmatic only for one pair of conjugate on-axis points. If the lens shown in Fig. 3.3 is to be stigmatic not only for the points, O and O' , but also for the points, P and P' , it must satisfy the *Herschel condition*

$$n dz \sin^2 \frac{u}{2} = n' dz' \sin^2 \frac{u'}{2}$$

Figure 3.3

Definition of Paraxial Lens



If the same lens is to be stigmatic at the off-axis conjugate points, Q and Q' , it must satisfy the *Abbe sine condition*

$$n dy \sin u = n' dy' \sin u'$$

Generally, these two conditions cannot be met exactly and simultaneously. However, if the angles u and u' are sufficiently small, and we can substitute the sine of the angle with the angle itself

$$\sin u \approx u \quad \text{and} \quad \sin u' \approx u'$$

then both the Herschel and Abbe sine condition are satisfied. We say that the lens works in the *paraxial region*, and it behaves like a perfect stigmatic lens. The other common definition of paraxial optics is that paraxial rays are rays “infinitely close to the optical axis.” This is a fine and correct definition; however, it can become difficult to understand when we consider tracing a paraxial ray through the edge of a lens system, a long way from the optical axis. This creates a dilemma since rays traced through the edge of the system are hardly infinitely close to the optical axis! This is why the first definition of paraxial optics, that is, using the small-angle approximation to the ray tracing equations, as would be the case for rays infinitely close to the optical axis, is easier to understand. Consider Fig. 3.4a, where we show how the rays are refracted at the interface between two optical media, according to Snell’s law.

$$n \sin \theta = n' \sin \theta'$$

Figure 3.4

A Real Ray Trace and a Paraxial Ray Trace through a Lens

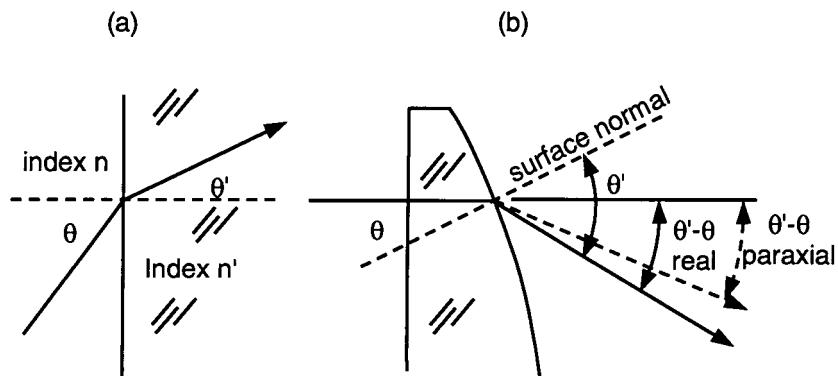


TABLE 3.1

Paraxial Approximation Versus Real Ray Angles of Refraction

θ	$\theta' - \theta$ real (degrees)	$\theta' - \theta$ paraxial (degrees)	Difference (degrees)
1	0.5001	0.5	0
10	5.0981	5	0.0981
20	10.8659	10	0.8659
30	18.5904	15	3.5904
40	34.6186	20	14.6186

In Table 3.1 we show just how a real ray, according to Snell's law, and a paraxial ray, using the small-angle approximation $\sin \theta \sim \theta$, refract or bend after refraction from a spherical surface at a glass-air interface (index of refraction of glass $n = 1.5$). These data use the nomenclature of Fig. 3.4*b*. Note that the difference in angle between the paraxial and the real rays define the resulting image blur. For angles of incidence θ of 10° or less we see that the real refracted ray is descending within 0.1° of the paraxial ray (0.0981° difference at a 10° angle of incidence). However, as the angles of incidence increase, the difference between the real and the paraxial descending angles increases quite significantly. This is where aberrations come from.

Along with this understanding, it is evident that in order to keep aberrations small, it is desirable if not mandatory to keep the angles of incidence as small as possible on the various surfaces within your system.

What Is Diffraction?

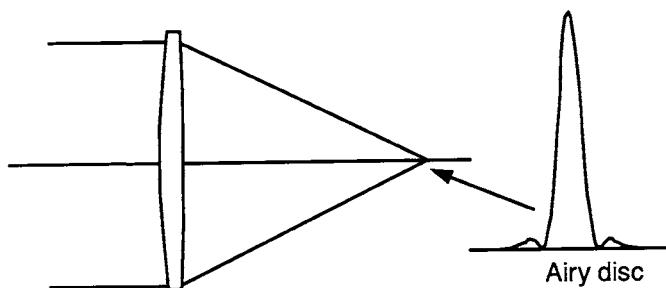
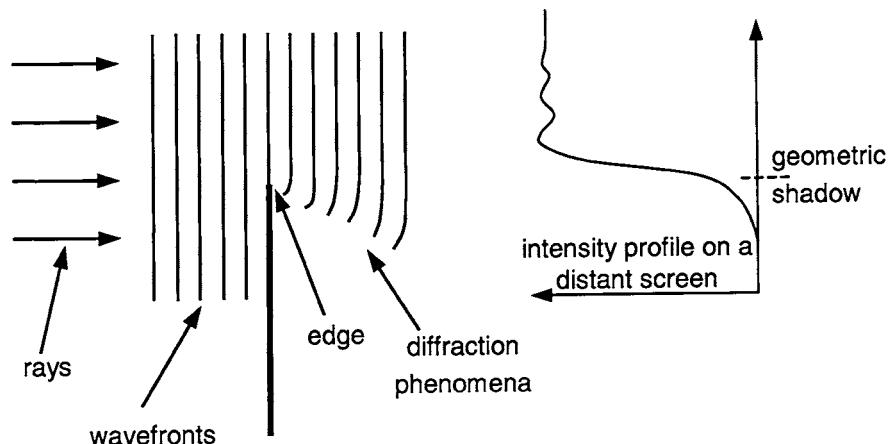
Diffraction is a phenomena or effect resulting from the interaction of light (which of course is electromagnetic radiation) with the sharp limiting edge or aperture of an optical system. While we could very easily fill the next few pages with integral signs and Bessel functions, it is not the intention of the authors to provide this level of detail. Rather, the following explanation is easy to follow and should provide a sufficient level of understanding of the causes of diffraction and resultant observable effects.

Imagine a swimming pool at 3 o'clock in the morning with no wind present; the water is like a sheet of glass. Imagine throwing a large rock

into one end of the pool. Water waves will emanate outward from where the rock has entered the water as concentric, ever-expanding circles. Before proceeding onward, note that the physics of water waves is virtually identical to the physics of electromagnetic radiation, and all of the derivations are quite analogous.

Now let us proceed to the other end of the swimming pool. If the pool is large enough, the water waves will be nearly straight and parallel to one another. In reality, of course, they are going to be curved and centered about the point where the rock entered the water. For this discussion, consider the water waves as straight. Let us now immerse a 1- by 2-m sheet of plywood partially into the pool, as shown at the top of Fig. 3.5. What you will see to a reasonable extent above the edge of the board at the top of Fig. 3.5 is that the water waves will continue to propagate left

Figure 3.5
Diffraction Effects



to right undisturbed. Below the edge where the major part of the board is located, you will see to the right of the board virtually no disturbance in the water. To the right of the intersection of the upper edge of the board and the water waves, you will see little curlicues traveling or emanating outward from the edge of the board. These curlicues are, in reality, *diffraction* of water waves. We sincerely hope that none of our readers would think there would be a sharp step in the water to the right of the edge... if you did think this was the case, we urge you to try this little experiment in your backyard swimming pool.

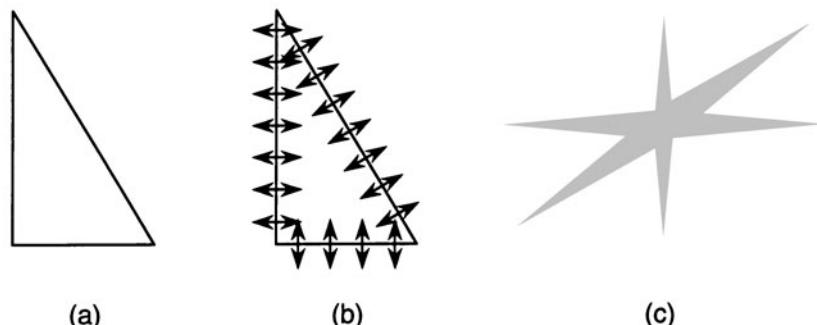
The peaks of the water waves are called *wavefronts*. Perpendicular to the wavefronts are the *rays*. While we rarely if ever talk about "water rays," we certainly do talk all the time about light rays. The important point here is that the rays are perpendicular or normal to the wavefronts and the wavefronts are perpendicular to the rays. Throughout your reading of this book, we would like you to understand the difference between ray optics and wave optics.

Now, back to our little example. If instead of our swimming pool example we were to have parallel or collimated light incident upon the edge of a razor blade, we would have diffraction of the electromagnetic radiation much in the same way as we showed diffraction of water waves. On a distant screen or card you will not see a very sharp edge or step function, but rather an intensity gradient with slight variations in intensity occurring in a similar fashion to the curlicues of the water waves.

The previous explanation represents our attempt to illustrate how diffraction occurs without the lengthy and messy mathematical derivations. If we now have a typical lens system, it is easy to understand that there will, by definition, be a limiting edge or aperture at which the light effectively stops or is blocked. This edge, which in many cases is the aperture stop of your system, wraps around the optical axis generally in a symmetrical, circular fashion, and the resulting diffraction pattern acquires a rotationally symmetric shape known as the "Airy disk."

It is important to note that diffraction occurs perpendicular to an edge. Since a circular aperture, in effect, wraps around in a full 360° , the resulting diffraction pattern (the Airy disk) is a rotationally symmetrical blur. However, if your aperture were a triangular shape as shown in Fig. 3.6a, the resulting diffraction pattern would be star shaped with three spikes as shown in Fig. 3.6c. The reason there are three notable spikes is that the diffraction spreading has occurred perpendicular to the three straight edges of the aperture, as shown in Fig. 3.6b. Note that the relative length of the spikes is proportional to the length of the edge.

Figure 3.6
Diffraction from a Triangular Aperture



Diffraction-Limited Performance

As discussed previously, if the geometrical aberrations are significantly smaller than the diffraction blur, the image is, in effect, well represented by the Airy disk. This form of optics is called *diffraction-limited optics*. Understanding the limits of diffraction-limited optics becomes extremely important, especially with today's extremely demanding levels of performance. Figure 3.7 shows two very important principles: (1) the physical diameter of the Airy disk and (2) the angular diameter or subtense of the Airy disk. It can be shown that:

$$\text{Physical diameter of the Airy disk} = 2.44 \lambda f/\text{number}$$

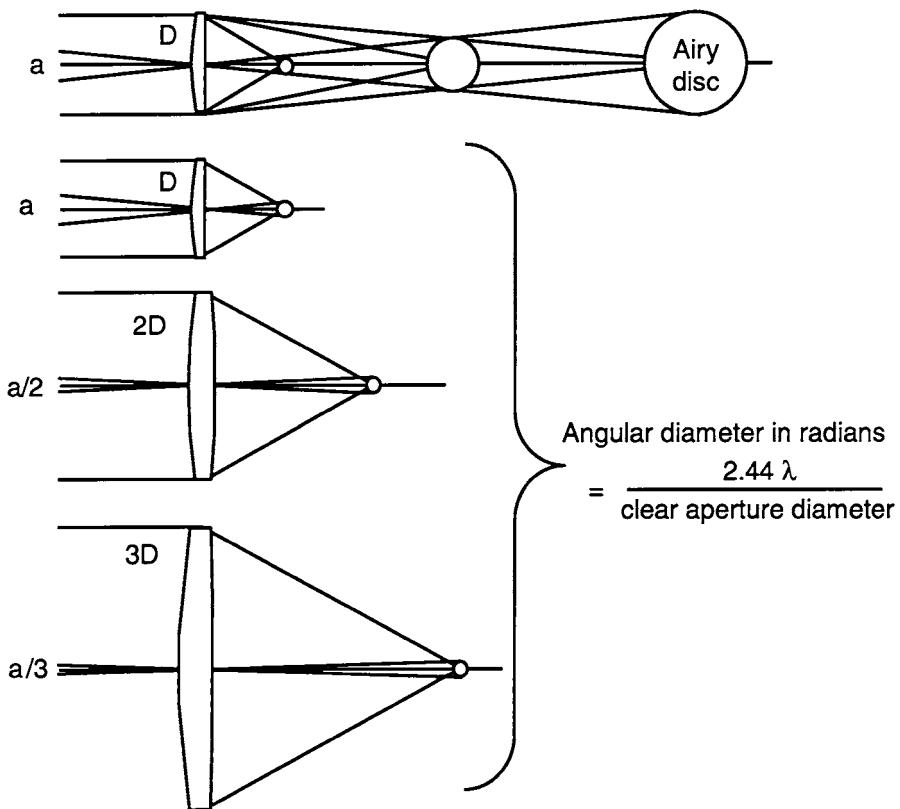
Shown in the top part of Fig. 3.7 are three different lenses, all of diameter D . One lens focuses images fairly close to the lens, the second has a somewhat longer focal length, and the third a still longer focal length. For these three lenses, all of which have the same entrance pupil diameter D , the $f/\#$ increases in proportion to the increase in focal length. From the equation, we see that the Airy disk increases in diameter in direct proportion to the $f/\#$ increase and thus in proportion to the focal length as well. A very useful rule of thumb to remember is: *The Airy disk diameter in the visible part of the spectrum is approximately equal to the $f/\#$ expressed in microns.*

This is easy to see if you consider a wavelength of $0.5 \mu\text{m}$, which would be approximately the center of the visible spectrum. In this case, the physical diameter of the Airy disk

$$D = 2.44 \times 0.5 f/\# = 1.22 f/\#$$

Figure 3.7

A Clarifying Illustration of "Diffraction-Limited" Imagery



which is approximately equal to the *f*/number itself expressed in microns!

We now show three separate lenses of diameter D , $2D$ and $3D$, all of identically the same *f*#. What this means is that the Airy disk or diffraction blur will be identical in all three cases. You can see quite clearly that for each lens the focal length increases in proportion to the increase in diameter (since the *f*# is identical). What this means is that the angular subtense of the Airy disk also decreases in proportion to the diameter or the increase in focal length. The resulting relationship becomes

$$\text{Angular diameter of the Airy disk} = \frac{2.44 \lambda}{\text{clear aperture diameter}}$$

The angular diameter is expressed in radians if the wavelength and the clear aperture diameter are in the same units.

Note that in all of the preceding discussion, the diameter of the Airy disk is assumed to be the diameter of the first dark ring in the diffraction pattern.

Derivation of System Specifications

There is a broad term “systems analysis” which generally refers to the task of deriving the basic optical system parameters based on the functional system performance requirements. We can apply what we learned earlier to perform a simple, yet noteworthy systems analysis example.

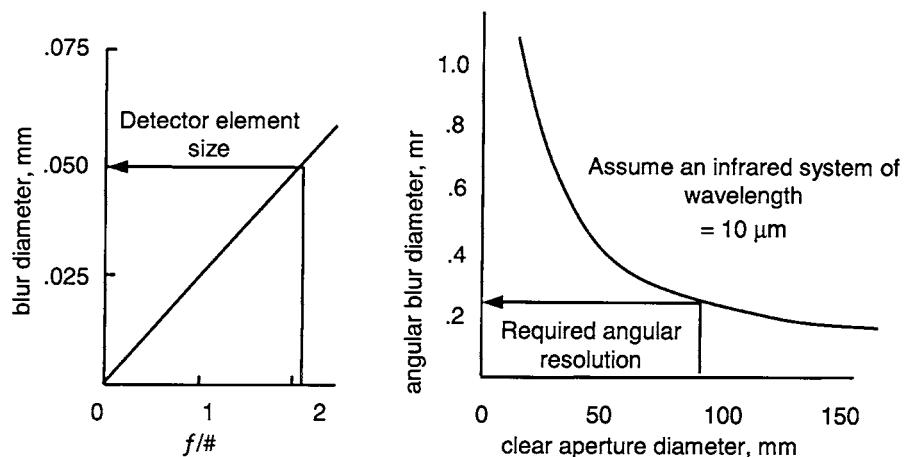
Consider, for example, an optical system used in the long-wave infrared (LWIR) which operates in the 8- to 12- μm spectral band. Our task is to derive the system $f/ \#$ and clear aperture diameter. Let us assume that the detector is mercury cadmium telluride or (HgCdTe) with a 50- μm pixel pitch and further assume that we need to resolve 0.25 mrad in object space. These values are typical for an LWIR system such as a forward looking infrared (FLIR).

Earlier in this chapter, we discussed that as a rule of thumb the smallest resolvable image blur should be matched to the pixel size of the detector (sensor), that is, smallest element size. Thus, we would require that the diffraction blur or Airy disk should be approximately the same diameter as our 50- μm pixel. Recall that the diameter of the Airy disk, $D = 2.44 \lambda f/\#$, and we can solve for the $f/\#$ to produce a 50- μm Airy disk diameter, and the result is $f/2.2$ at $\lambda = 10 \mu\text{m}$. Before we continue, it is interesting to note that for diffraction-limited optics, an $f/2.2$ system that is 6 mm in diameter will have exactly the same diffraction blur diameter as an $f/2.2$ system that is 3 m in diameter, and that is 50 μm !

For a given diffraction blur diameter, as the focal length increases, the angular subtense of the Airy disk decreases proportionally. We can use the relationship that the angular diameter of the Airy disk = $2.44 \lambda / (\text{clear aperture diameter})$ to solve for the clear aperture diameter required so that the 50- μm Airy disk subtends 0.25 mrad in object space, and the result is a 100-mm-diameter clear aperture.

Figure 3.8 shows parametrically how the $f/\#$ and clear aperture diameters relate to the diffraction-limited image blur or Airy disk diameter and the angular subtense of the diffraction blur. This illustrates how we can quickly and easily take the most basic system functional requirements and derive the system $f/\#$ and clear aperture diameters. Do keep

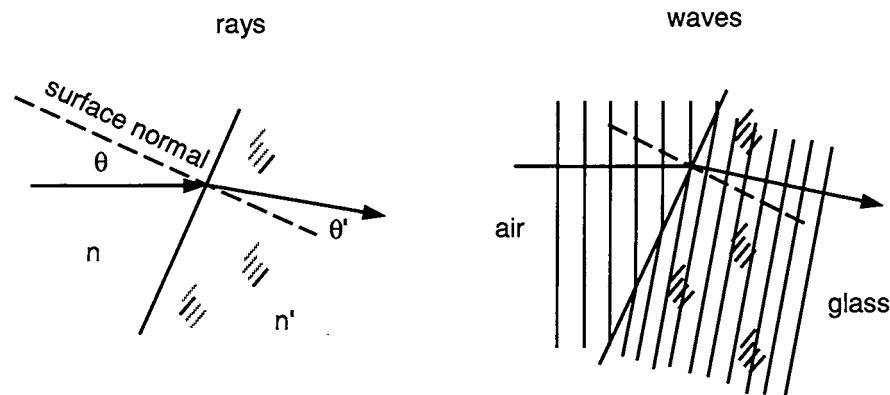
Figure 3.8
Example of Systems Analysis



in mind that this assumes diffraction-limited optics. Further, it is based on the criteria that the Airy disk is matched to the pixel pitch. These are generally good assumptions to work with, and as your system needs become better understood, you may need to revise the results.

As we begin to learn more about image formation, it is important to understand just how light bends or refracts when passing through an air-glass or glass-air interface. As shown in Fig. 3.9, the *refractive index of a material* = n , where n is the ratio of the velocity of light in a vacuum to the velocity of light in the denser material such as glass. Since the light or electromagnetic radiation slows down in the denser material, the refractive index is always greater than unity. For optical glass the refractive index ranges from about $n = 1.5$ to $n = 1.85$.

Figure 3.9
Bending of Light at an Optical Surface



According to the Snell's law, $n \sin \theta = n' \sin \theta'$. With air on the input side of the interface, the equation reduces to $\sin \theta = n' \sin \theta'$. For small angles the equation further reduces to $\theta = n'\theta'$. We will use this result later!

We discussed earlier how light could be represented by either rays or wavefronts, where the two are orthogonal to one another. We will be using both representations throughout this book and we hope that you will become "bilingual" or fluent with both representations. To help understand these concepts, we show in Fig. 3.9 how a light ray, as well as a series of wavefronts, is incident on an air-glass interface and how the light bends or refracts. From ray optics, we can simply use Snell's law to determine the angle of refraction. Consider how we represent the same thing using wave optics. The wavefronts are traveling from left to right, with their peaks separated by the wavelength of light. As the wavefront enters the denser medium such as glass, its velocity is reduced by $1/n$, with the result being that the wavefronts are closer together. There is a fundamental law of physics, which says that the wavefronts must be continuous at the interface between the media. Considering the velocity reduction along with the wavefront continuity requirement, we can see how the entire wavefront is rotated around in a clockwise direction as it proceeds into the denser medium. Interestingly, you can use this construction to rederive Snell's law!

The Concept of Optical Path Difference

Optical Path Difference (OPD) and the Rayleigh Criteria

OPD is an extremely useful measure of the performance of an imaging optical system. If the wavefronts proceeding to a given point image are spherical, concentric, and centered at the point image for a given field of view, then the imagery will be geometrically perfect, or *diffraction limited*. As shown earlier, the image will then be a perfect Airy disk. This is, in effect, the reverse of our earlier example where we threw a rock into a pool of water to illustrate the wave nature of light and diffraction. If we think of the water waves traveling in reverse to where the rock entered the water, we will emulate light imaging to a point image. By definition, the wavefronts will be perfectly spherical, concentric, and centered where the rock entered the water. Recall also that rays are perpendicular to the wavefronts. It is thus clear that if the wavefronts are spherical, concentric, and centered at a point in the image, then the rays will all come to that same point as defined by the center of curvature of the wavefronts. As we learned earlier, diffraction at the limiting edge of the pupil will create an Airy disk, which is the reason why we do not have a perfect point image.

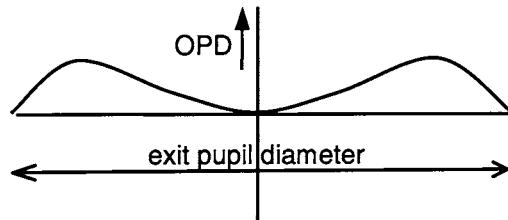
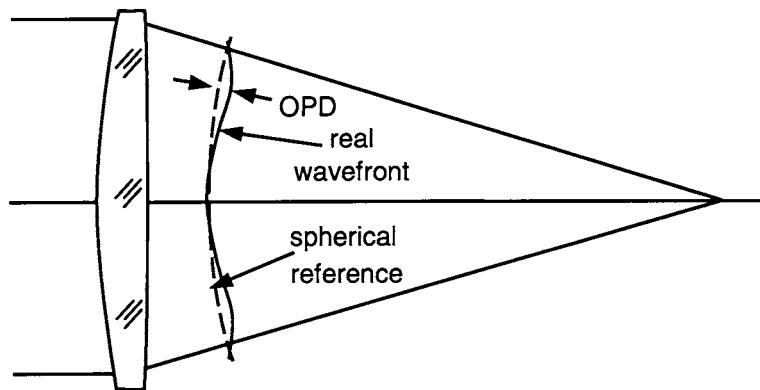
Consider Fig. 4.1 where we show a hypothetical lens with a perfectly spherical reference wavefront and a real wavefront. The real wavefront departs from sphericity due to aberrations induced by the lens. The *optical path difference* is the difference between the real wavefront and a spherical reference wavefront, which is usually selected to be a near best fit to the aberrated wavefront.

One of the reasons the OPD is so valuable a parameter is evident from the Rayleigh criteria. Lord Rayleigh (real name William Strutt, a Nobel Prize winner for discovering the gas argon) showed that:

An optical instrument would not fall seriously short of the performance possible with an absolutely perfect system if the distance between the longest and shortest paths leading to a selected focus did not exceed one-quarter of a wavelength.

What the Rayleigh criteria says is that if the OPD is less than or equal to one-quarter of a wave (one-quarter of the wavelength of the light),

Figure 4.1
Optical Path Difference (OPD)



then the performance will be almost indistinguishable from perfect. If this is the case, then the imagery of a point object will be very nearly a perfect Airy disk. This is a *very* useful tool, and as will become evident, its validity is quite broadly applicable. It is important to note, however, that it is not 100% infallible, and should only be used as a guide or relative measure of the level of optical performance.

Figure 4.2 shows the appearance of the image of a point source, which is known as a *point-spread function* (PSF), for optical path differences of 0 waves (a perfect Airy disk), 0.25 wave, 0.5 wave, and 1.0 wave. Note that

Figure 4.2

Image of a Point Source with Different Amounts of Peak-to-Valley Optical Path Difference Due to Coma

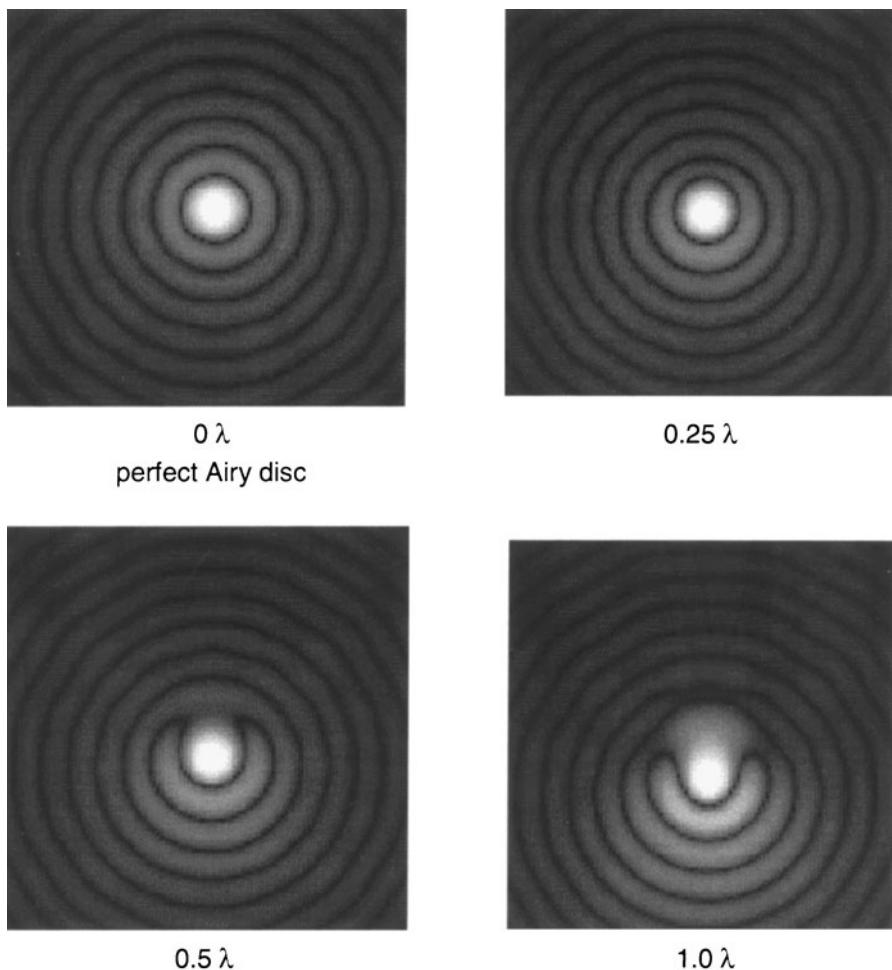
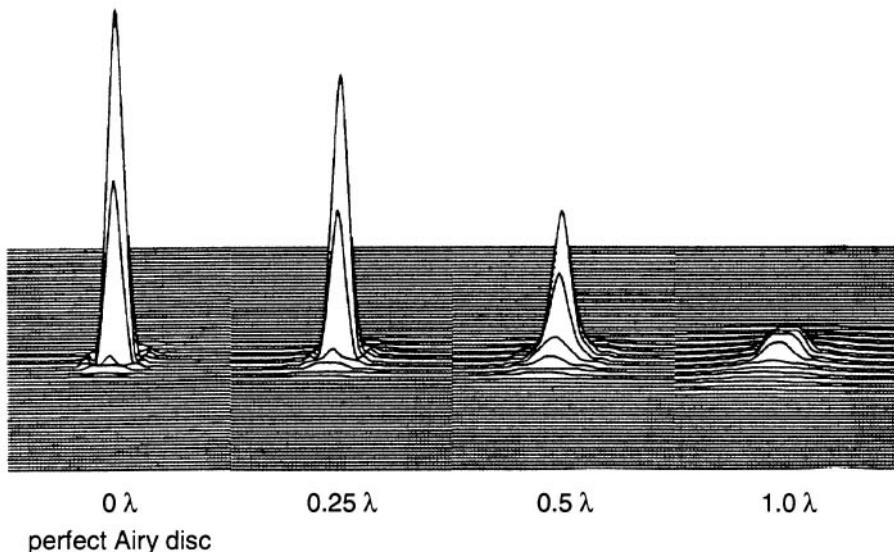


Figure 4.3

Image of a Point Source with Different Amounts of Peak-to-Valley Optical Path Difference Due to Spherical Aberration

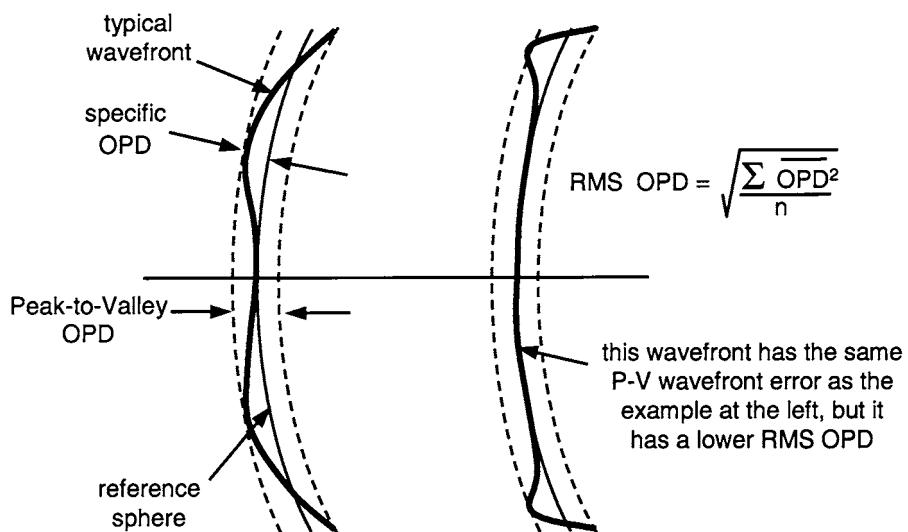


the 0.25 wave imagery is qualitatively nearly indistinguishable from the perfect Airy disk. The character of the central maximum is maintained, and the first bright ring is fully intact. As soon as we go to 0.5 wave and above, the imagery is clearly degraded from perfect. Figure 4.3 shows perspective views of the point spread function for the same values of the OPD; however, these data are for spherical aberration rather than coma. Here, too, we can conclude that the 0.25 wave imagery is nearly indistinguishable from perfect. We do see a drop in peak intensity; however, the overall character of the pattern is very similar to the perfect Airy disk.

Peak-to-Valley and RMS Wavefront Error

The OPD as shown here is known as *peak-to-valley* (P-V) optical path difference. Peak-to-valley is the total difference between the portion of the wavefront closest to the image (leading, or ahead of the reference wavefront) and the farthest lagging portion of the wavefront (lagging, or behind the reference wavefront). Figure 4.4 shows this as the separation between the two dashed reference spheres.

Figure 4.4
Peak-to-Valley and
rms Wavefront Error



There is another term, and that is *rms wavefront error*. The definition of rms wavefront error is shown in Fig. 4.4 as the square root of the sum of the squares of the OPDs as measured from a best-fit reference spherical wavefront over the total wavefront area. The rms wavefront error represents more of an averaging over the wavefront than the P-V wavefront error. The example shown on the right-hand side of Fig. 4.4 has the same P-V OPD as the left-hand side; however, the rms would be lower. This is because most of the wavefront error, or wavefront distortion, is at the outer periphery of the aperture, and over most of the area of the wavefront, the wavefront is nearly perfect.

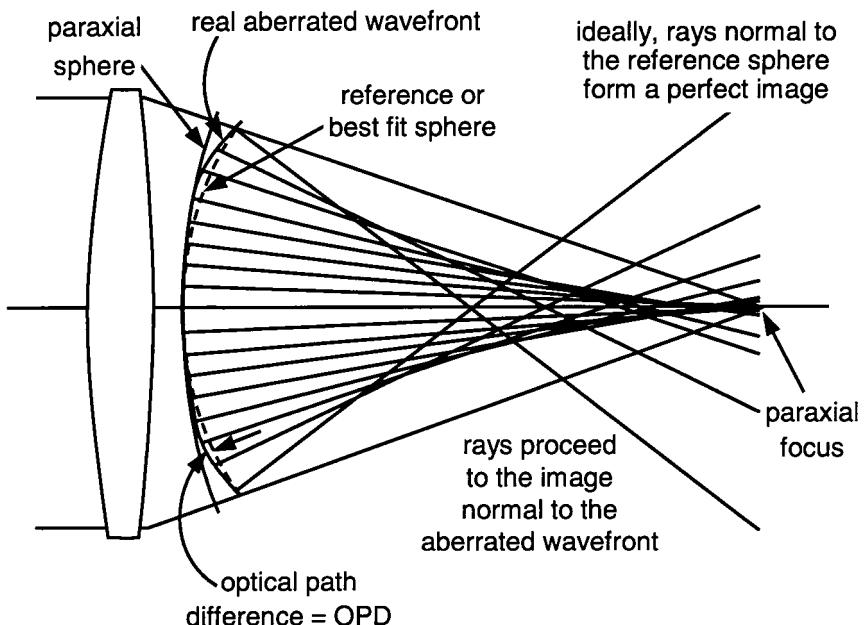
Consider, for example, a large telescope mirror 3 m in diameter. In order to assure near-diffraction-limited performance, let us assume that the P-V wavefront error is specified as 0.125 wave on the surface. This is the P-V departure from the ideal or perfect surface profile. In reflection, the wavefront departure will be double this value, or 0.25 wave which just meets the Rayleigh criteria. Now let's further assume that the optical shop produces a mirror which has a P-V surface departure from the nominal of 0.02 wave, with the exception of a small depression the size of a pencil eraser 0.5 wave deep. The mirror is clearly out of spec as the reflected wavefront will have a P-V error of 1.0 wave, which is four times the Rayleigh criteria. However, the area of this small depression in the surface would be 0.0004% of the total mirror area, an almost negligible amount. This will have virtually zero effect on the optical performance

of our telescope, and if the scattering from such an error were of concern, we could simply paint the 6-mm-diameter depression with a flat black paint. While there will still be some scattering from the mirror/paint interface, this, too, will be extremely negligible in all but the most demanding applications (such as with space telescopes).

The rms wavefront error typically ranges from approximately one-fifth to one-third of the P-V error. This ratio is highly dependent on the correlation of the wavefront, where the correlation is the inverse number of bumps over the surface. For a given number of bumps, a lower correlation has greater surface slope errors and conversely if we assume a ratio of $5\times$ between P-V and rms, the Rayleigh criteria of 0.25 wave P-V equates to 0.05 wave rms.

Figure 4.5 shows an exaggerated illustration of just how an aberrated wavefront proceeds to an image. This figure reminds us of several key points such as the fact that rays are perpendicular to the wavefront. The peak-to-valley OPD is the maximum deviation from the real wavefront and a spherical reference wavefront, which best fits the real wavefront. While the figure is quite exaggerated, it is drawn to scale and the various factors we have learned about all apply.

Figure 4.5
OPD Showing Wavefronts and Ray Paths



The Wave Aberration Polynomial

The optical path difference, or the wave aberration function, can be mathematically expressed in the form of a polynomial for rotationally symmetric optical systems.

A single ray proceeding from a given point in the object through an optical system is defined by the coordinates in the object plane and its coordinates in the pupil of the system. The wave aberration function can be expressed as a Taylor expansion polynomial in field and pupil coordinates. The wave aberration polynomial can be simplified by using the symmetrical properties of the optical system. In its final form the wave aberration polynomial has two quadratic terms which are not the intrinsic aberrations—they present a focal shift, five terms of the fourth power of the field and pupil coordinates which are primary aberrations, and sixth, eighth, and tenth, etc., power terms which are higher-order aberrations.

In order to obtain the coefficients in the wave aberration polynomial, it is sufficient to trace a small number of rays and then fit the data to the polynomial. To obtain the higher-order aberrations, it is necessary to do finite ray tracing, but the primary ray aberrations can be calculated by a paraxial ray trace. In optical systems with moderate to small apertures and fields, primary aberrations dominate. The wave aberration polynomial, W or OPD, is of the form

$$W = W_{020} r^2 + W_{111} h r \cos \theta + W_{040} r^4 + W_{131} h r^3 \cos \theta + W_{222} h r^2 (\cos \theta)^2 + W_{220} h^2 r^2 + W_{311} h^3 r \cos \theta + \dots \text{(higher-order terms)}$$

where h is the height of the object and r and θ are polar ray coordinates in the pupil (see Fig. 4.6).

It can be shown that the ray coordinates in the image plane relative to the perfect image coordinates are proportional to the partial derivatives of the wave aberration polynomial, that is,

$$\partial y' \sim -\frac{\partial W}{\partial y} \quad \partial x' \sim -\frac{\partial W}{\partial x}$$

This means that if the OPD or the wave aberration polynomial is known, the ray intersections in the image plane or spot diagrams can be easily calculated. The exponent of the pupil radius term is higher by one in the wave aberration polynomial than in the ray-intercept equations. Thus, for example, third-order spherical aberration affects the

Figure 4.6

Nomenclature for
Wave Aberration
Polynomial

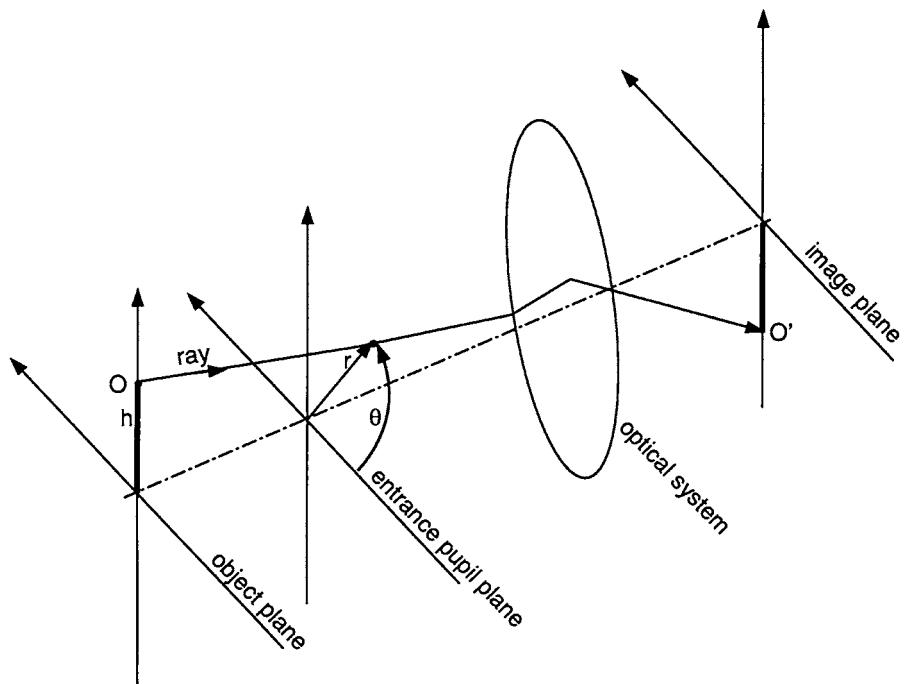


image blur diameter in proportion to the cube of the radius of the pupil, whereas the optical path difference is proportional to the fourth power of the pupil radius.

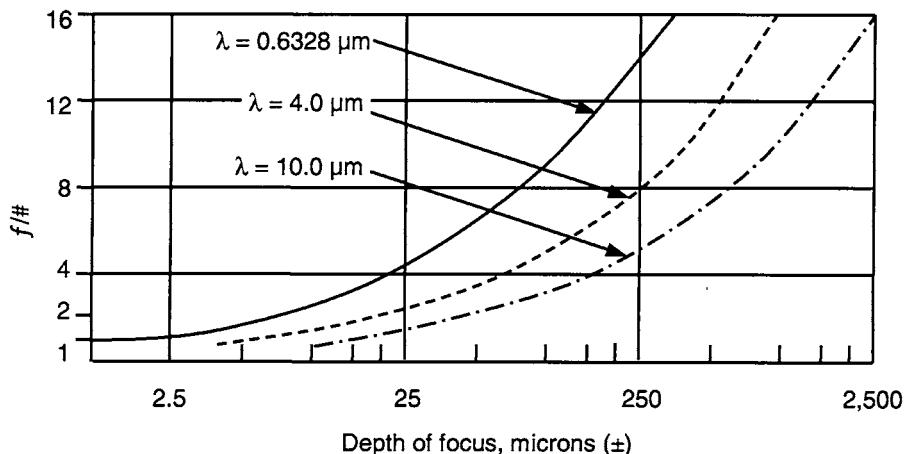
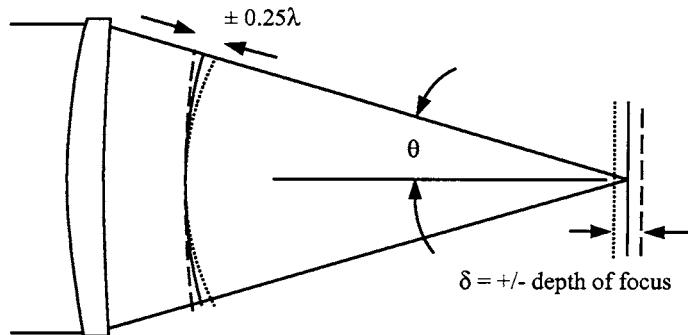
Depth of Focus

As we now know, if the optical path difference is less than or equal to $1/4\lambda$, our system meets the Rayleigh criteria and the system imagery is nearly indistinguishable from perfect. This result can be effectively used to determine just how much defocus is tolerable to maintain diffraction-limited performance.

Consider an otherwise perfect optical system, as shown in Fig. 4.7. The solid line in the upper part of Fig. 4.7 represents the nominally perfect spherical wavefront proceeding to the nominal image plane. If we now locate a compass point displaced fore and aft of the nominal image plane and draw two circles which touch the nominal wavefront on the

Figure 4.7

Depth of Focus



optical axis, these circles will depart from the nominal wavefront along the limiting marginal ray by an amount which is, in effect, the optical path difference. We now adjust the compass point until this displacement from the nominal wavefront is ± 0.25 wave. This yields the image plane locations which correspond to one-quarter wave of defocus. The depth of focus which corresponds to an OPD of $\pm 1/4\lambda$ is

$$\delta = \pm\lambda/(2 n \sin^2 \theta) = \pm 2 \lambda (f/ \#)^2$$

An extremely useful rule of thumb is that the depth of focus in the visible is approximately $\pm(f/\#)^2$, in micrometers. Thus, for an $f/4$ lens in the visible the depth of focus is approximately $\pm 16 \mu\text{m}$. For an $f/2$ system

the depth of focus is approximately $\pm 4 \mu\text{m}$, and so on. In the lower portion of Fig. 4.7, we show the depth of focus for systems in the visible, the medium-wave infrared (3 to 5 μm), and the long-wave infrared (8 to 12 μm), respectively. This is shown as a function of $f/\#$. It will become very apparent that as the $f/\#$ and wavelength increase, the depth of focus increases as well. This increase is linear with wavelength and quadratic with $f/\#$.

Do keep in mind that this assumes an otherwise perfect system. If your lens system has some inherent aberrations and/or wavefront errors due to design or manufacturing errors, then it will not be nominally perfect to begin with and you may not be able to allow a full one-quarter wave of depth of focus in image location or defocus before you degrade the performance by the quarter-wave limit.

There is another term, “depth of field,” which is often confused with depth of focus. Both terms are defined here, which should dispel any further confusion.

Depth of focus. The amount of image defocus which corresponds to being out of focus by one-quarter wave. This means that the optical path difference between the real wavefront leaving the exit pupil at its outer periphery and a reference wavefront centered at the nominal image plane is one-quarter of the wavelength of light.

Depth of field. This is a term used mostly in photography. What it means is that if you focus a camera at a given distance or range, how much further from the camera and closer to the camera than this distance will objects be in acceptable focus. This is analogous to depth of focus; however, it is not as stringent, and it is directly related to how acceptable the image looks to the eye.

Review of Specific Geometrical Aberrations and How to Get Rid of Them

As discussed in Chap. 3, aberrations are the failure of the optical system to produce a perfect or point image from a point object. The geometry of focusing light using spherical surfaces is simply not perfect, and spherical surfaces are used primarily due to their inherent ease of manufacturing. Many lenses can be ground and polished at the same time, as was shown in Fig. 3.2. Lenses are blocked together on the rotating part of the machine called a “block.” The top part, which is called a tool, has the desired radius of curvature, and it moves back and forth as the block rotates, forming spherical surfaces of the same radius on all lenses.

As was discussed, paraxial optics applies Snell’s law using the small angle approximation where the sine of the angles of incidence on surfaces is equal to the angle, in radians. In paraxial optics, there are no aberrations whatsoever, and by definition, the image of a point object is a perfect point image. Aberrations occur because in a real system the angles of incidence are nearly always so large that the paraxial

approximation is invalid and this causes the rays not to converge to a single point image.

As will be discussed in Chap. 7, the use of nonspherical, or aspheric, surfaces can often help significantly in minimizing, if not eliminating, aberrations. It is important to note that the use of aspherics does not automatically guarantee that the aberrations will be zero; in fact, for the most part, this will not happen. Their use is yet another technique for minimizing and balancing aberrations. There are techniques for manufacturing aspheric surfaces or aspheric lenses such as injection molding of plastic lenses, compression molding of glass, or diamond-turning aspheric surfaces in plastic, some crystals, or metal. Aspheric surfaces are used for additional aberration correction, but for the most part, spherical surfaces are used in optical systems. Aspheric optical components are often expensive, such as diamond-turned surfaces and glass-molded lenses, and not sufficiently accurate or unstable with a change of temperature such as plastic lenses.

There is a class of small lenses used in optical storage applications such as CD read lenses where aspherics are mandatory. These lenses are about the diameter of an aspirin tablet and are compression molded or manufactured by other techniques. In addition, many of today's digital cameras contain very small lenses (less than 6 to 8 mm in diameter) and glass aspherics are becoming more common in these application areas.

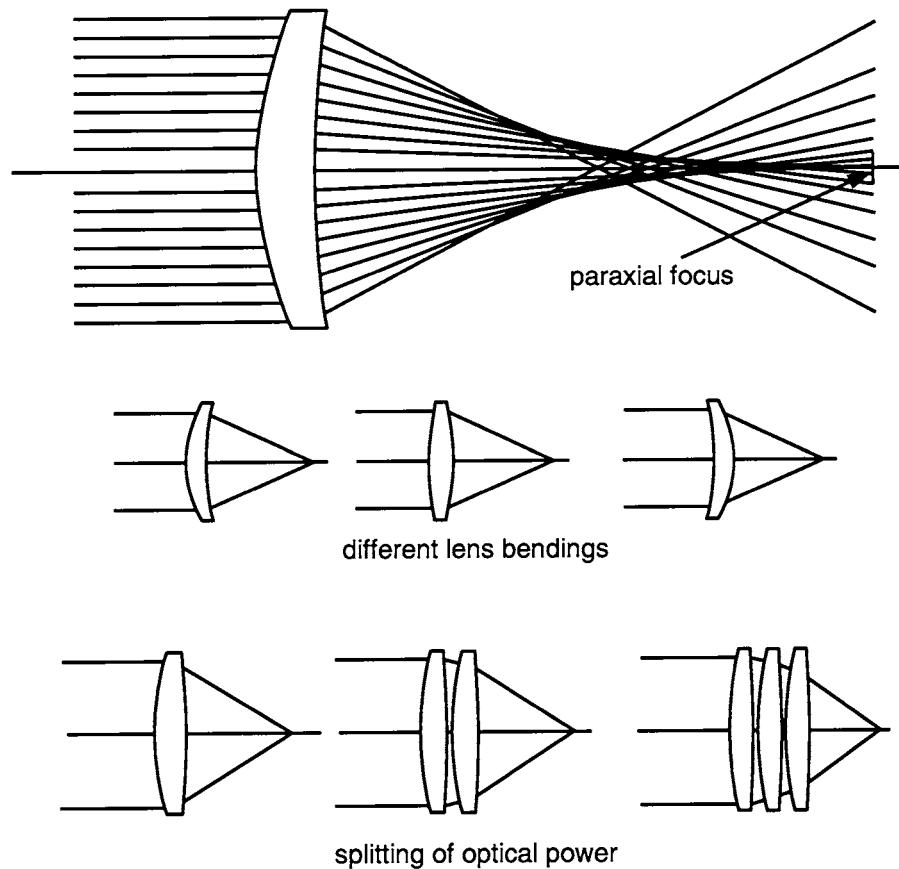
The index of refraction of the glass and other transmitting materials is used for making lenses changes with the wavelength of light, a phenomena called *dispersion*. The result is aberrations which change as a function of the wavelength. These aberrations are called *chromatic aberrations*. The image of a point is a superposition of the images for the entire wavelength band or spectral range, each of them blurred with the presence of monochromatic aberrations.

With a well-chosen combination of optical parameters such as lens shapes, number of optical elements, and different optical materials, aberrations in real optical systems with large ray angles can be reduced to a minimum or may be able to be eliminated to the level of the diffraction limit.

Spherical Aberration

If light is incident on the single lens shown in Fig. 5.1, rays that are infinitely close to the optical axis will come to focus at the paraxial image position. As the ray height above the optical axis at the lens

Figure 5.1
Spherical Aberration



increases, the rays in image space cross the axis or focus closer and closer to the lens. This variation of focus position with aperture is called *spherical aberration*.

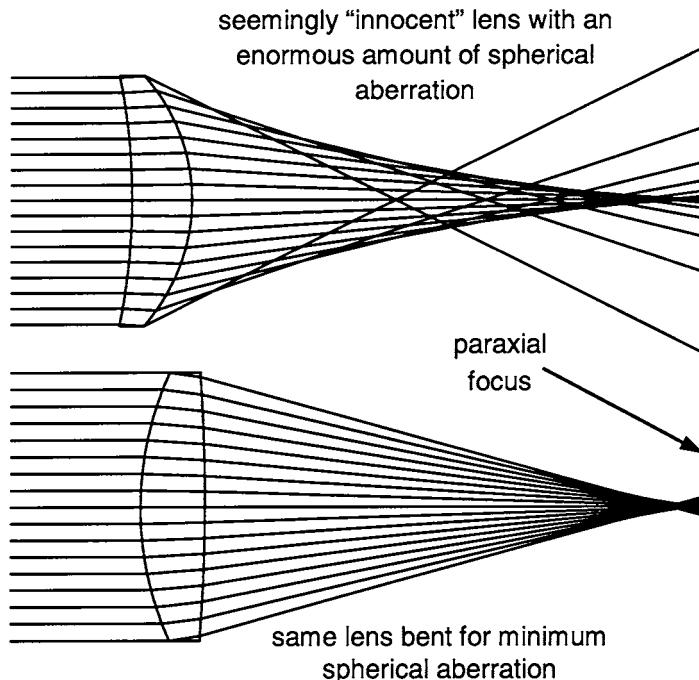
The magnitude of this spherical aberration depends on the height of the ray in the entrance pupil. The amount of spherical aberration is proportional to the cube of ray height incident onto the lens. If the spherical aberration is measured along the optical axis, it is called *longitudinal spherical aberration*. More often, it is measured as a lateral or transverse aberration, and it represents the image blur radius. For a given focal length lens, a lens with twice the diameter will have eight times larger image blur. For a given focal length and aperture of a single lens, spherical aberration is a function of the object distance and bending (shape) of the lens.

Also shown in Fig. 5.1 are lenses of different bending. The meaning of the term “lens bending” is that the focal length and hence the power of the lens is maintained while changing the radii of both surfaces. This would be the same as physically bending a lens made of flexible plastic. Spherical aberration is highly dependent on the relative lens bending, as will be discussed later.

Another powerful method of controlling spherical aberration is by splitting the optical power into more elements, as shown in the lower portion of Fig. 5.1. By splitting the optical power among several elements, the angles of incidence on each surface can be reduced, resulting in reduced aberrations. As we learned earlier in Chap. 3, reducing the angle of incidence results in a smaller deviation between paraxial rays and real rays, and hence reduced aberrations.

Consider Fig. 5.2 where we show a single $f/2$ lens element with an enormous amount of spherical aberration. The lower part of Fig. 5.2 also shows an $f/2$ lens; however, in this case the lens is bent for minimum spherical aberration.

Figure 5.2
Spherical Aberration
as a Function of Lens
Bending



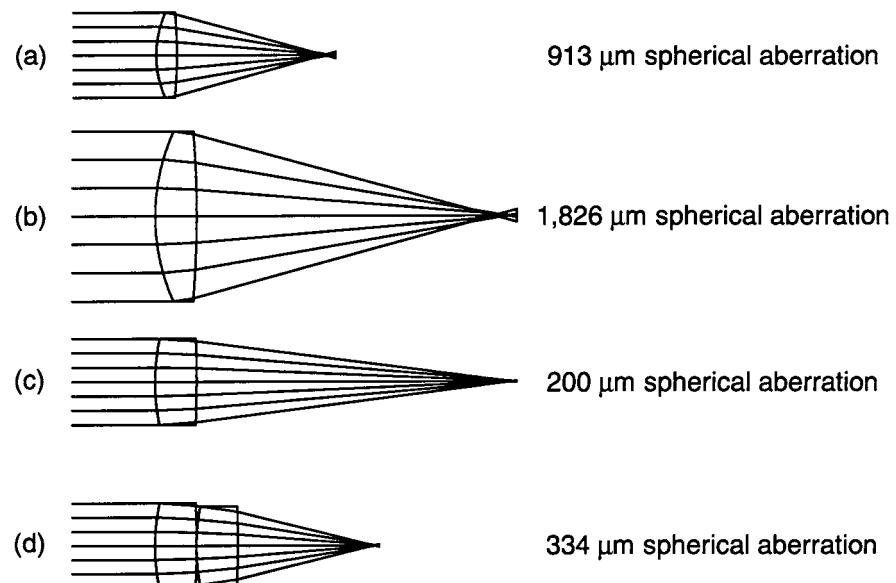
When the object point is at a finite distance, the shape of the lens changes for minimum spherical aberration. In a symmetrical case, when the distance of the object point from the lens is the same as the distance of the image, an equiconvex lens is the bending which produces minimum spherical aberration.

Let's look further into reducing the spherical aberration by splitting a lens into several elements. The resulting lens will perform the same function, keeping the total optical power of the elements the same as of the original lens.

We will demonstrate in a simple way how spherical aberration can be reduced by a factor of 2, if the lens is split into two lenses. We will do this in several logical steps:

1. The first step is to start with a lens bent for minimum spherical aberration, and this is shown in Fig. 5.3a for a 25-mm-diameter $f/2$ lens of BK7 glass. The residual spherical aberration is 913 μm from real ray tracing.
2. We now scale the lens up by a factor of 2, as shown in Fig. 5.3b. The focal length of our new lens is twice as large, the diameter is twice as large, and the spherical aberration is also twice as large. Note that when we scale a lens, all parameters with units of length scale by the same factor such as the radii and thickness.

Figure 5.3
Splitting Optical Power to Reduce Spherical Aberration

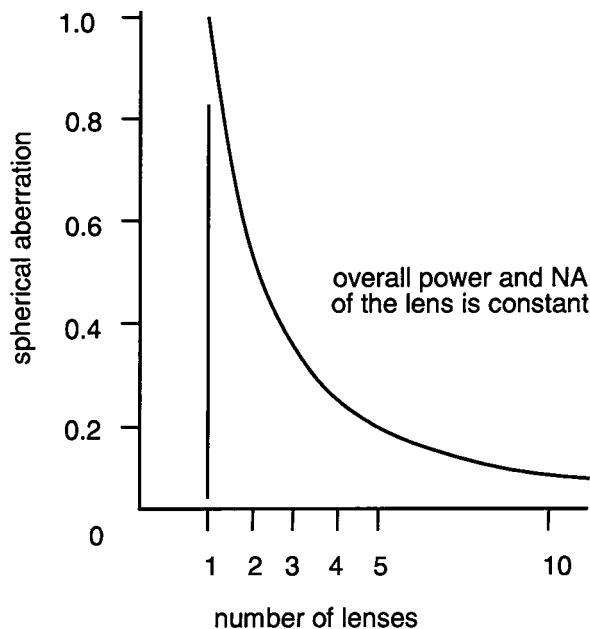


The refractive index is unitless, and thus remains unchanged. The spherical aberration is now doubled to 1826 μm .

3. We now reduce the aperture of this scaled-up lens by a factor of 2, as shown in Fig. 5.3c. The spherical aberration reduces by the cube of the aperture, which means by a factor of 8, which is 228 μm . Real ray tracing gives 200 μm , which is quite close. Now we have a lens with approximately four times less spherical aberration than the starting point. This new lens has the same aperture as the starting lens, but its focal length is twice as large.
4. Now we add one more identical lens of the same power (same focal length), as in Fig. 5.3d. The spherical aberration is doubled (approximately), but it is twice as small as the aberration of the starting lens. The real ray tracing shows that our final solution has 334 μm of spherical aberration, which is 36% of the starting value with a single element of the same focal length.

The new configuration consists of two lenses performing the same function as the single starting lens, but having one-half of the spherical aberration. The theoretical result of splitting a single lens into multiple lenses is shown in Fig. 5.4. This result shows that if we split an element

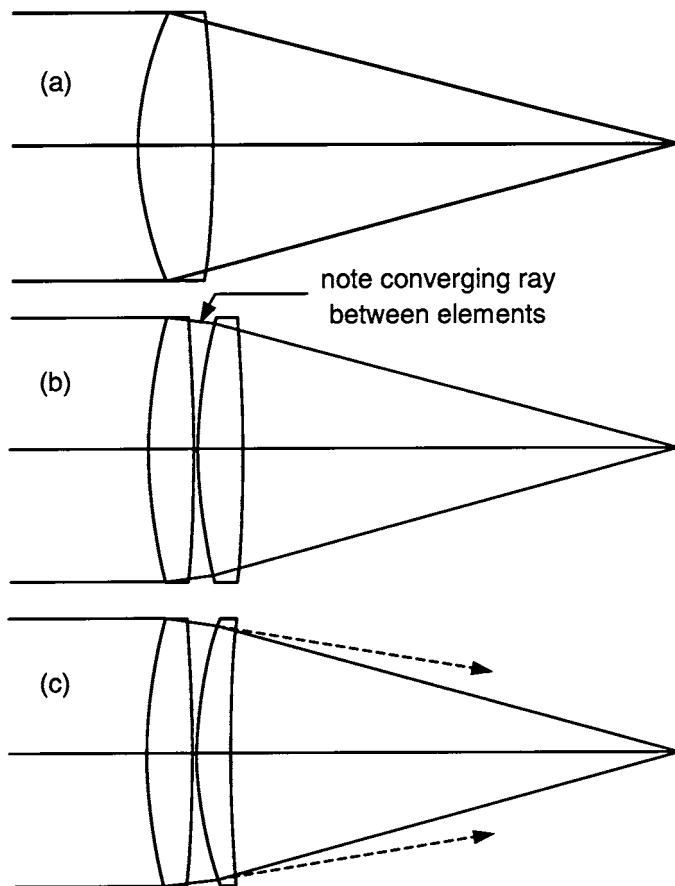
Figure 5.4
Spherical Aberration
as a Function of
Number of Elements



into four to five elements, the spherical aberration will reduce to about 10 to 15% of the single-element starting point. As it turns out, when we split power in a real lens, the results are significantly better. This is because the light exiting the first element will be converging, and if the second element is now bent for minimum spherical aberration based on converging incident light, the resulting spherical aberration is reduced even further. This way, by introducing even more optical elements, the spherical aberration can be reduced significantly. Figure 5.5 shows the situation. Figure 5.5a represents a single element bent for minimum spherical aberration. Figure 5.5b shows two identical elements as derived previously. Note that the light between the two elements is converging as it enters the second element. Figure 5.5c shows how the second element

Figure 5.5

Illustration of How to Achieve a Further Reduction of Spherical Aberration

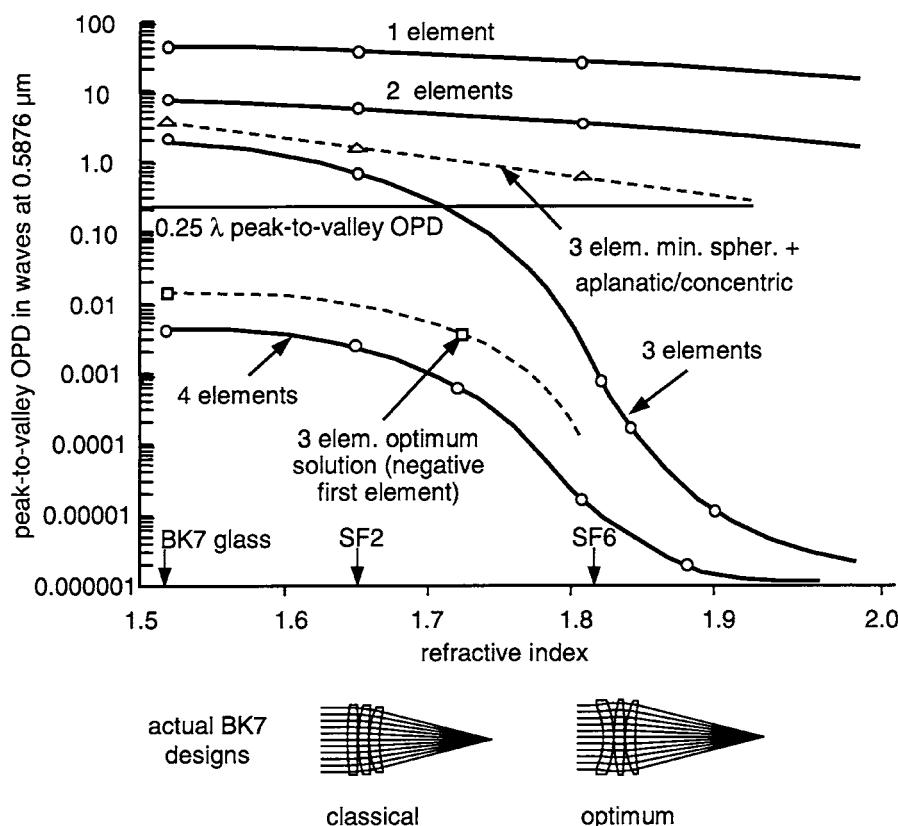


can more optimally “curl” or bend more strongly so as to minimize the angles of incidence onto its surfaces, thereby reducing the spherical aberration from the design in Fig. 5.5*b* with the identical elements.

It is instructive to consider the design of an $f/2$, 100-mm focal length lens for minimum spherical aberration with one, two, three, and four components, and we will do this for glasses with refractive indices ranging from 1.52 to 1.95. We will now plot the peak-to-valley optical path difference for all of these cases in Fig. 5.6.

The results are quite dramatic. Note that for a single $f/2$ element with a 100-mm focal length of BK7 glass (refractive index 1.517) the spherical aberration is approximately 40 to 50 waves P-V. Splitting the element into two elements reduces the OPD to about 6 to 8 waves, and splitting it into three elements further reduces it to about 2 waves. And four elements results in about 0.004 wave, a significant reduction. There

Figure 5.6
Spherical Aberration
as a Function of
Number of Elements
and Refractive Index



is a further reduction in OPD as the refractive index is increased, especially for three and four elements where nearly six orders of magnitude reduction in OPD is achieved by simply increasing the refractive index from 1.5 to 1.9!

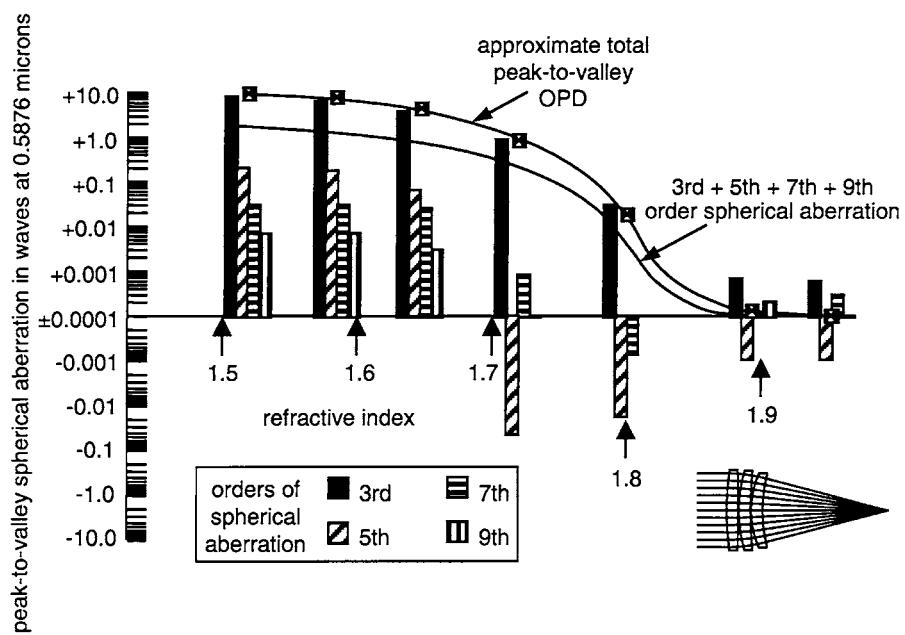
This is for what we have termed the “classical” solution. This is where each element is bent somewhat more than its predecessor in order to minimize the angles of incidence and thus the overall spherical aberration. As we will see later, there is a configuration which yields an even better solution, and we call this the “optimum” configuration. It is characterized by a negatively powered meniscus first element, and it yields two orders of magnitude less OPD than the classical solution, even at the lower refractive index region.

Note that the preceding parametric analysis is based on monochromatic light and was computed only on axis. While this is somewhat of an idealized situation, the insight we have gained into aberration reduction is of major significance and it further enhances our understanding of aberrations and where they come from. However, from the analysis thus far we really do not know just why such a dramatic reduction in the aberration is achieved.

Spherical aberration terms in the wave aberration polynomial are the fourth, sixth, eighth, etc., order in terms of the pupil radius. The exponent of the pupil radius term is larger by one in the wave aberration polynomial than in the ray-intercept equations. When we talk about spherical aberration image blur size in the image plane, we talk about third, fifth, seventh, ninth, etc., order in terms of the pupil radius. Let us again look at three component lenses optimized for the smallest spherical aberration, and compare the lenses from low-index glass $n = 1.5$, and then increase the index up to $n = 2$. The spherical aberration as a function of the index of refraction is shown in Fig. 5.7. The contribution to the third-, fifth-, seventh-, and ninth-order spherical aberration is shown for refractive indices ranging from 1.5 to 2.0. Generally, lower orders of aberration have higher values, and they are predominant in the polynomial. As the index increases to somewhere around $n = 1.7$, the fifth-order spherical aberration changes sign and starts to balance the third-order aberration, so that the overall spherical aberration has a significant drop. Although the spherical aberration changes a lot with the change of the index of the components, there is only an imperceptible change in the shape of the lenses. The surfaces become a little shallower, but the overall shape of the lens remains the same.

Figure 5.7

Third-, Fifth-, Seventh-, and Ninth-Order Spherical Aberration Versus Refractive Index for Three Elements



As a final illustration of what is happening, consider Fig. 5.8 where we show the classical solutions for refractive indices from approximately 1.5 to 1.7. The graphical data is the deviation of the wavefront from perfect as a function of the normalized pupil radius. Note that the OPD decreases from about 2 waves P-V to about 0.25 wave. Figure 5.9 shows the data for refractive indices ranging from 1.8 to 1.95, and we see that the OPD reduces from 0.002 wave to several ten-thousandths of a wave.

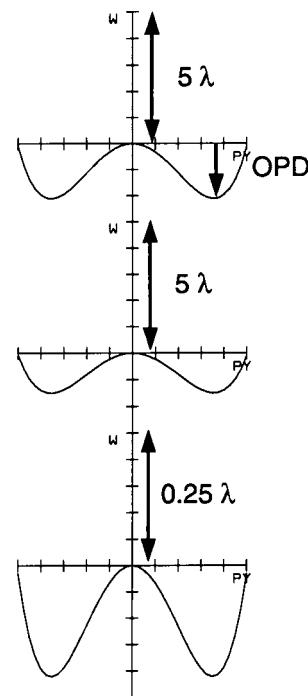
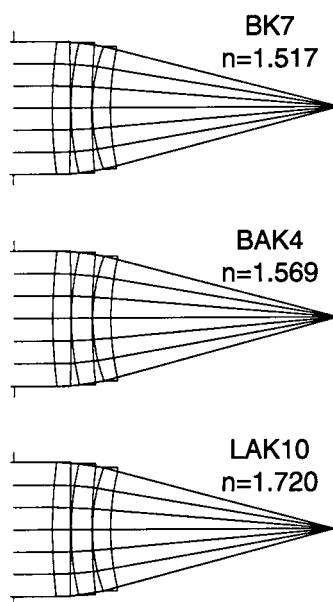
In all of these examples the relative shapes of the elements has remained nearly constant.

We noted earlier that there is a more optimum solution, and we show the three-element “classical” and optimum solutions along with the plot of optical path difference in Fig. 5.10. We are able to reduce the P-V OPD from 2 waves to less than 0.007 wave by changing the configuration. Both of these configurations use BK7 glass. The reduction in aberration is due to balancing of the fifth-order spherical aberration against the third-order, as described previously.

We have shown in the above material the different orders of spherical aberration, and how we can effectively and efficiently reduce the aberration by splitting the optical power into multiple elements, with the reduction of the angles of incidence serving as the primary means for

Figure 5.8

Lens Configuration and Plot of Optical Path Difference for Optimized Lenses of Refractive Index 1.517 to 1.720



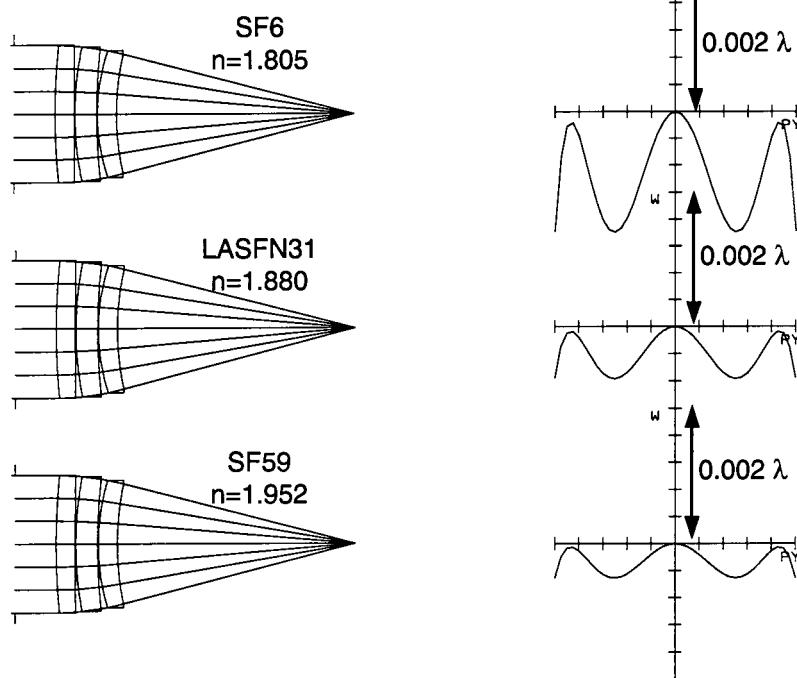
lower aberration. The lower the angle of incidence on a surface, the closer we are to the paraxial region where $n \times \theta = n' \times \theta'$ and the aberrations approach zero. There is a very interesting and useful tool called a "Pagel diagram" as shown in Fig. 5.11. The Pagel diagram is a plot, typically in the form of a bar chart showing the contribution of each surface within a lens system to the third order spherical aberration (other primary aberrations for a system with a finite field of view). Note how in this case (the classical solution) the contribution from the six surfaces are approximately +0.007, +0.007, +0.004, +0.004, -0.004, and +0.001, respectively. The sum of these contributions is +0.019, as shown.

It is shown in Fig. 5.12 how we can achieve a significantly higher level of performance by changing the first element to a negatively powered meniscus shape. Indeed, the negative contribution of the first surface balances nearly perfectly the positive contributions of the other five surfaces, for a net spherical aberration of nearly zero.

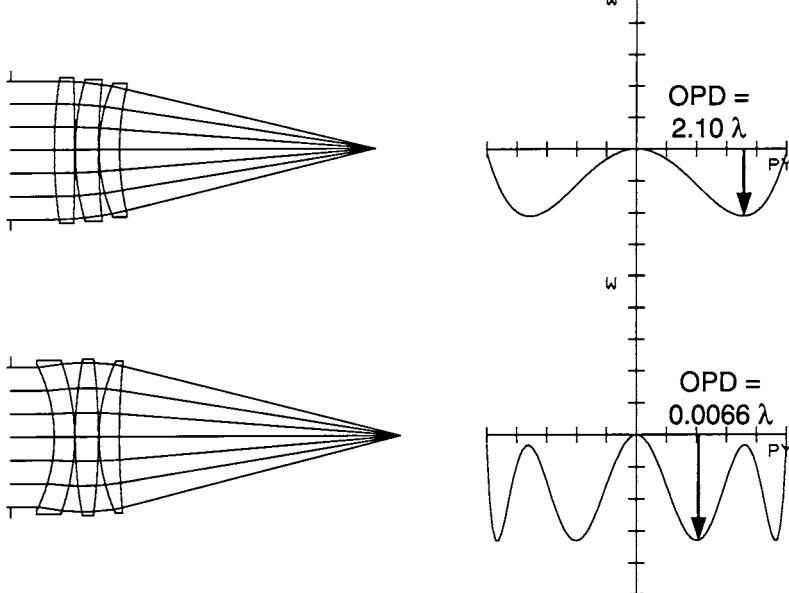
We have discussed orders of aberration in several contexts thus far. By carefully evaluating the plot of optical path difference, you can actually see visually the different orders. Consider Fig. 5.13 where we show again the

Figure 5.9

Lens Configuration and Plot of Optical Path Difference for Optimized Lenses of Refractive Index 1.805 to 1.952

**Figure 5.10**

"Classical" and Optimum Solutions with Three Elements for BK7 Glass



Review of Specific Geometrical Aberrations

Figure 5.11
Pagel Diagram of
3 Element "Classical"
Solution Showing
Surface Contributions
to Spherical Aberration

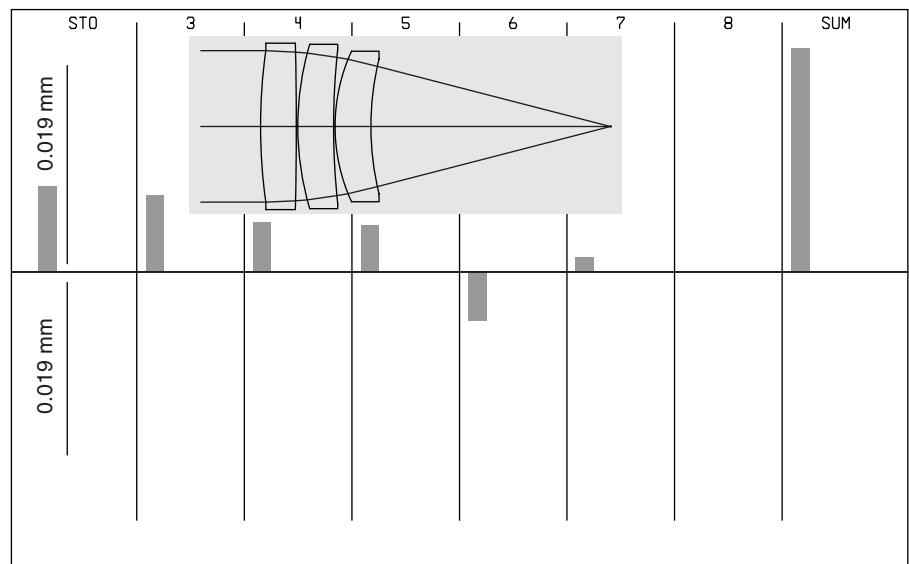


Figure 5.12
Pagel Diagram of
Optimum Solution

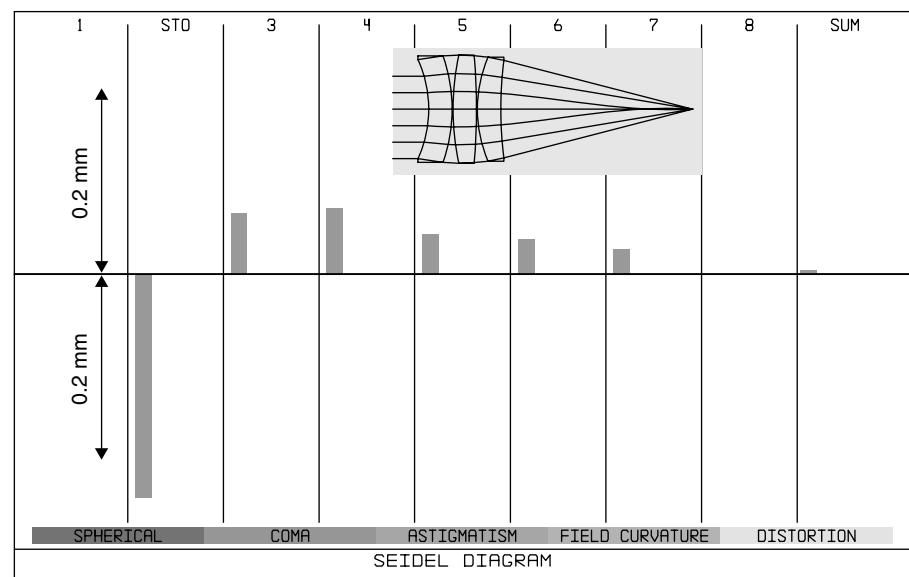
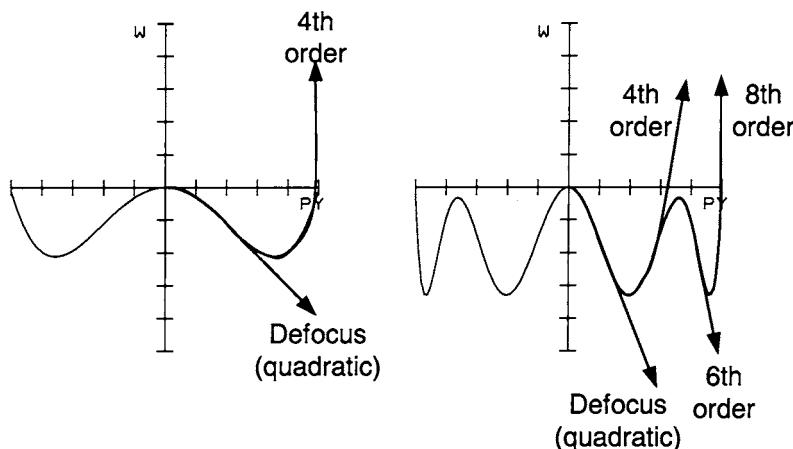


Figure 5.13
Illustration of the
Orders of Wavefront
Aberration



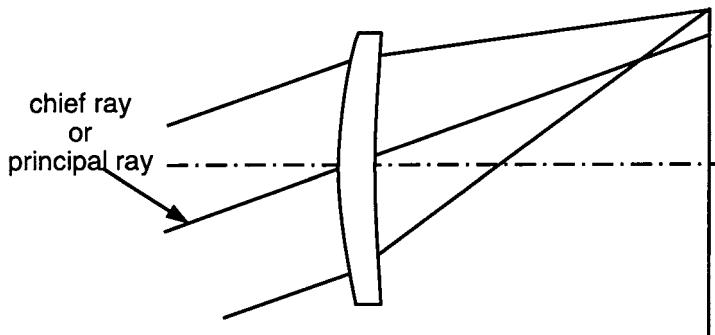
plots of OPD for the “classical” and optimum designs just discussed. However, here we show each of the orders of spherical aberration. Recall that the exponent is one higher than the transverse ray aberration polynomial, so the third-order spherical aberration is proportional to the fourth-order in OPD and so on. Of course, focus shift is thus quadratic in OPD as we would expect. Note that each time we see an inflection in the data this is equivalent to another order being added. There will almost, by definition, be higher orders than those shown; however, these data show clearly the presence of the different orders and how they tend to balance each other.

Coma

When we move away from the optical axis in field of view, the image of a point becomes nonrotationally symmetric. In Fig. 5.14 parallel rays come from an infinitely distant point which does not lie on the optical axis of the lens. They enter the lens at an angle, and they are focused by the lens to a certain height from the optical axis defined by the field angle and the focal length of the lens. If the lens itself limits the bundles of rays from different points in the field, we say that the aperture stop is located on the lens. Rays that go through the center of the aperture stop are called *chief rays*. There is only one chief ray for each point in the object.

Rays that go through the aperture stop and lie in the plane of the drawing are called *meridional rays*. Rays which do not lie in the meridional

Figure 5.14
Coma



plane are called *skew rays*. A plane perpendicular to the drawing in which lies the chief ray is called the *sagittal plane*. The meridional and sagittal planes have one common ray, the chief ray.

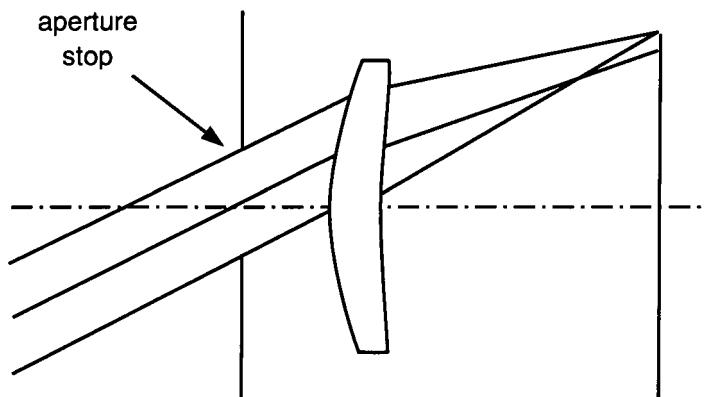
In an optical system *coma* is defined as the variation of magnification with aperture. Rays that transmit through the lens through different portions of the aperture stop cross the image plane at different heights from the optical axis. In the case of the single positive lens shown in Fig. 5.14, a ray passing through the top and bottom edge portions of the lens converges to a point in the image plane which is further from the optical axis than the point of convergence of other skew rays.

The shape of the image of a point as formed by a system with coma has the shape of a comet. The height of the image is usually defined by the position of the chief ray on the image plane. In the presence of coma, most of the light energy is concentrated in the vicinity of the chief ray. Coma is linearly proportional to the field of view and proportional to the square of the aperture.

When the aperture stop is not on the lens, moving the position of the stop can control the coma. Having greater ray symmetry on the way through the lens about the aperture stop reduces the coma. Figures 5.15 and 5.16 show the aperture stop to the left and the right of the lens, respectively, and it is clear that the off-axis ray bundles have a higher degree of symmetry and hence significantly reduced coma when the stop is to the right of the lens. This is due to the greater symmetry on the first surface, which results in reduced angles of incidence and hence reduced aberration.

We can best understand coma by looking at Fig. 5.17, which shows the cause of coma. In the top part of the Fig. 5.17, we show a collimated bundle of light incident obliquely onto a convex surface (we will only consider here the first surface of the lens). Note that the entire bundle is displaced

Figure 5.15
Coma with Stop in Front of Lens



from the normal to the surface, which is shown as a dashed line. Note also that the upper ray is incident onto the lens surface at a very high angle of incidence with respect to the surface normal, and as we know, this results in a significant ray bending or angle of refraction. The angles of incidence decrease rapidly as we transition to the lower portion of the ray bundle. The coma formation in this situation is quite evident.

Consider now the lower portion of Fig. 5.17, where we show a light bundle whose central or chief ray is normal to the surface. Now we have greatly reduced angles of incidence and, furthermore, the upper and lower limiting rays are symmetrical with each other. The net effect is that there is no coma contribution from this surface whatsoever, and the residual aberration is the same as spherical aberration.

Figure 5.16
Reduced Coma with Stop Aft of Lens

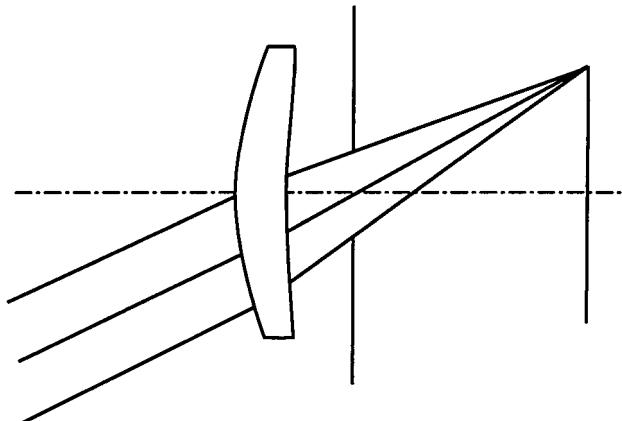
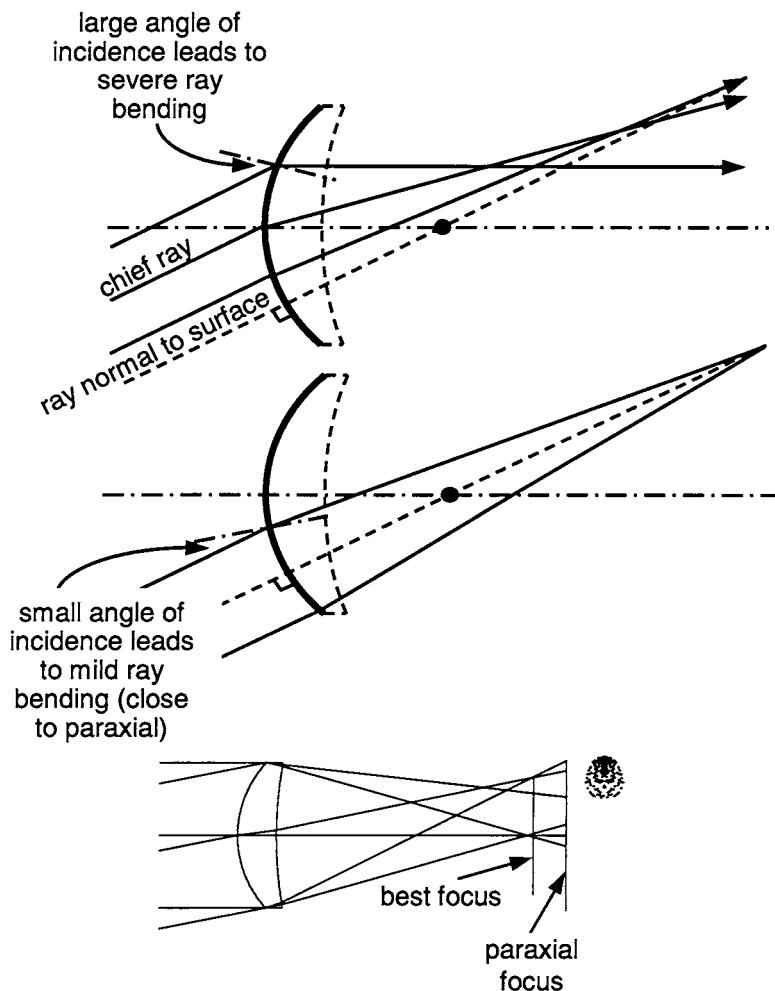


Figure 5.17
Where Does Coma Come From?

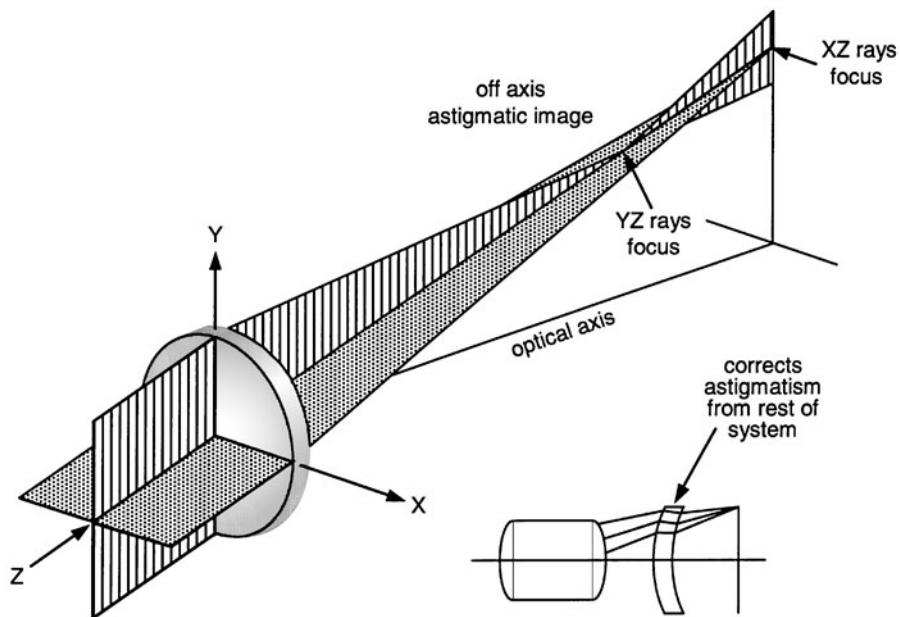


Astigmatism

In the presence of astigmatism, rays in the meridional and sagittal planes are not focused at the same distance from the lens. An astigmatic image formed by a positive lens is shown in Fig. 5.18.

Rays in the meridional plane focus along the line that is perpendicular to the meridional plane. Rays in the sagittal plane are focused further away from the lens, along the line perpendicular to the sagittal plane. Between the astigmatic foci, the image of a point is blurred. It takes the

Figure 5.18
Astigmatism



shape of an ellipse or circle. The smallest size of the image blur is half-way between two astigmatic foci when it is circular. Astigmatism is linearly proportional to the lens aperture and to the square of the field angle.

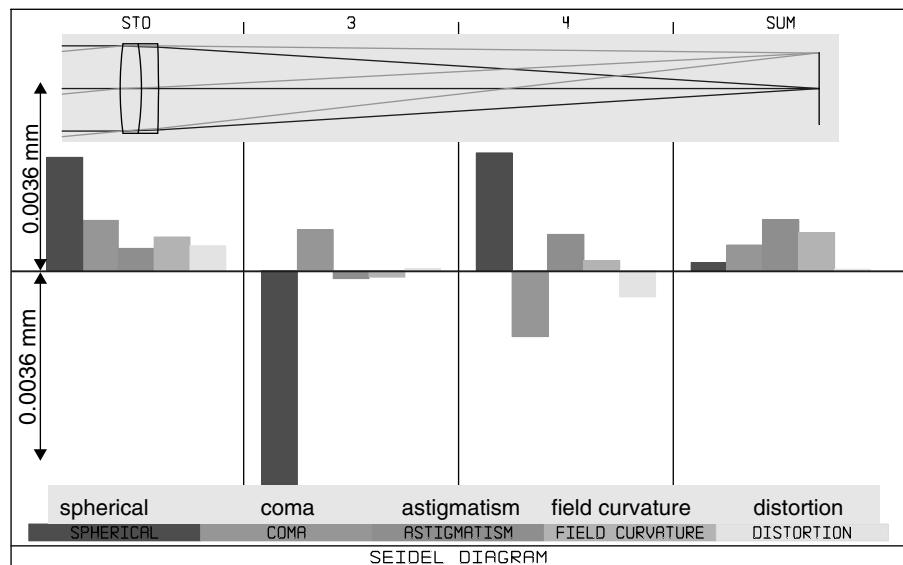
Astigmatism can be controlled by changing the shape of the lens and its distance from the aperture stop, which limits the size and position of the bundle of rays passing through the lens.

A tilted plate in a converging cone of light introduces astigmatism. A weak meniscus lens close to the image plane acts similar to a tilted plate with a tilt angle which changes from zero on axis to a certain angle at the edge of the field. This way, astigmatism created by the meniscus can partly or completely cancel the astigmatism of the rest of the optical system. Also shown in Fig. 5.18 is an example showing just how this works.

To show clearly how this works, Fig. 5.19 shows a Pagel diagram of an achromatic doublet with a small finite field of view. The astigmatism is the largest residual aberration. We now will add a weak meniscus element and reoptimize the design, with the result shown in Fig. 5.20. It is clear that the astigmatism at the image is virtually zero with the balancing of the aberration handled mostly by the meniscus element. Remember that this element form is analogous to a varying tilted plate in a converging cone as described previously.

Figure 5.19

Pagel Diagram of Achromatic Doublet



Where does astigmatism come from? An oblique cone of light incident on a lens is shown in Fig. 5.21. Assume that this element is immersed somewhere in the middle of an optical system. The area or footprint on the surface of the light cone shown extends over more of the surface in the y or tangential direction than in the x or sagittal

Figure 5.20

Pagel Diagram of Field Corrected Achromatic Doublet

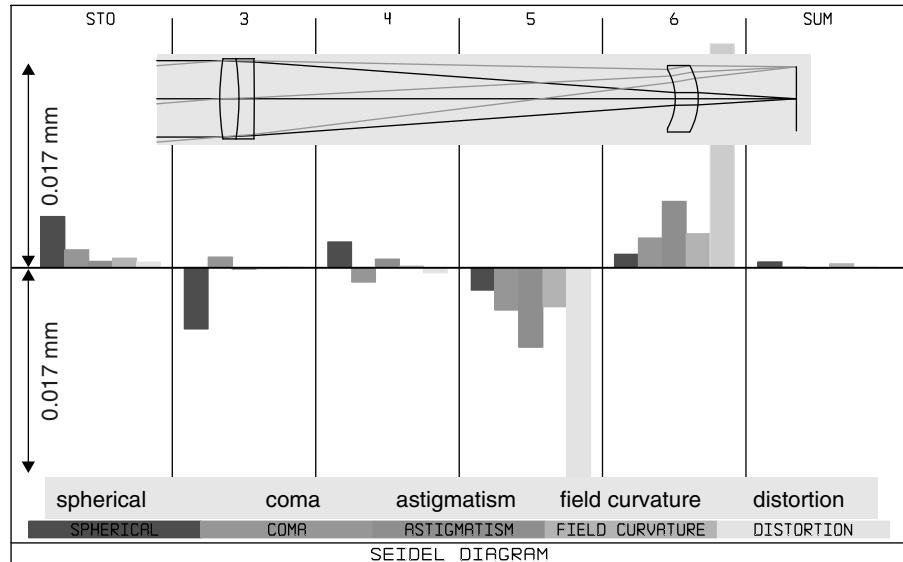
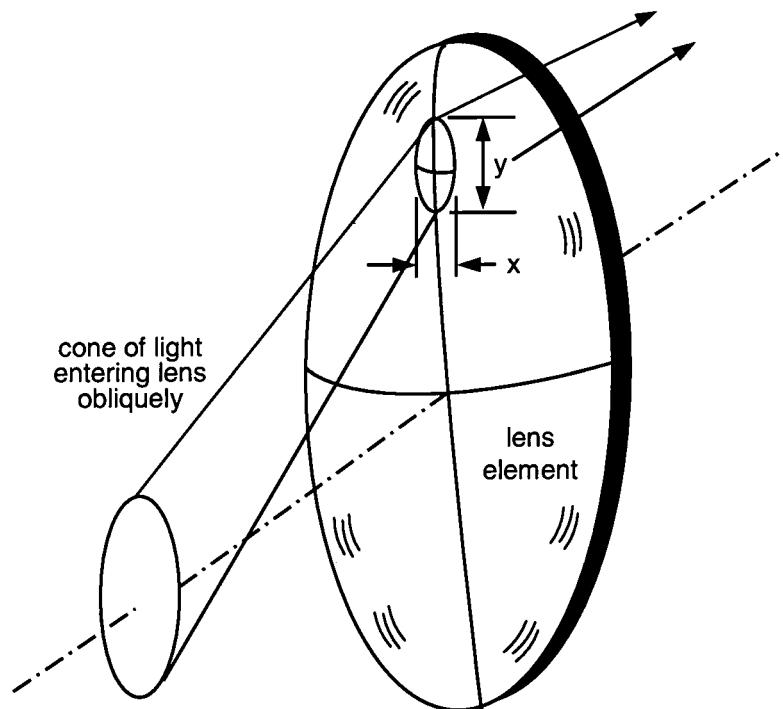


Figure 5.21
Where Does Astigmatism Come From?



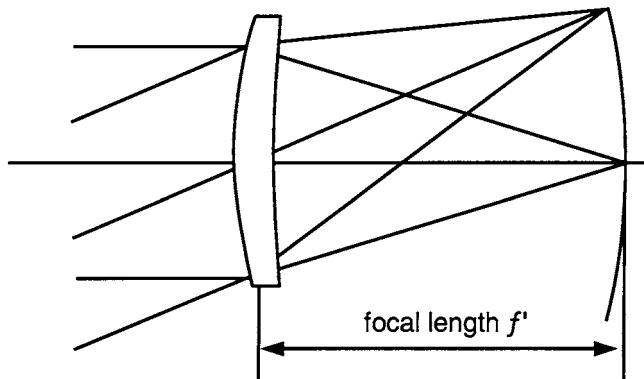
direction. Recall that the rate of change of slope of a sphere is constant everywhere on the sphere. Thus, the extreme tangential rays see a greater slope change on the surface than the extreme sagittal rays and are hence refracted at greater angles. This causes the tangential ray fan to focus closer to the lens than the sagittal ray fan, and this is astigmatism. As the surface is spherical, we will also find in many cases an off-axis form of spherical aberration called “tangential oblique spherical aberration” which is introduced in the tangential direction.

Field Curvature and the Role of Field Lenses

A positive lens forms an image on a curved surface, as shown in Fig. 5.22. In the absence of astigmatism, a surface on which the image is formed is called the Petzval surface. If a lens has no astigmatism, the sagittal and

Figure 5.22

Field Curvature



tangential images coincide with each other, and they coincide with the Petzval surface.

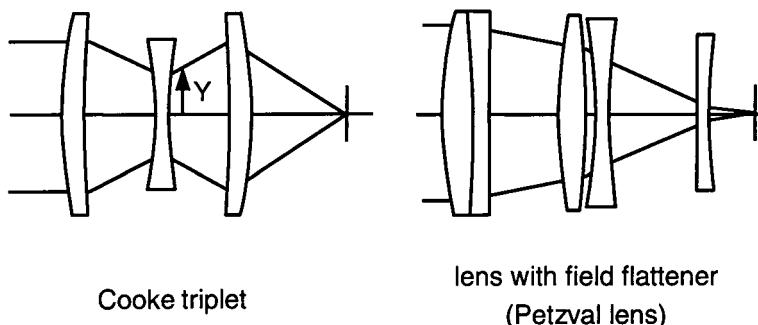
In the presence of astigmatism, both sagittal and tangential image planes are closer to the lens than the Petzval surface, and the tangential image is three times further from the Petzval surface than the sagittal image. The curvature of the Petzval image is inversely proportional to the product of the index of refraction of the lens and its focal length. If there are many components in the optical system, the resulting Petzval curvature is a sum of Petzval curvature contributions from all lenses.

We know that with a 35-mm camera, we can take nice sharp photographs using a flat film. Which method is then used in designing a lens to get a flat image plane? Since the contribution a lens element makes to the Petzval sum is proportional to its optical power, simply splitting of the elements will not change the field curvature. However, positively and negatively powered components can be combined to reduce the field curvature to zero. When negatively powered lenses are added to the system, the resulting power is also reduced.

Fortunately, there is a solution to this problem. The contribution a lens element makes to the system power is proportional to the product of its power and the height of the marginal ray which is the ray going through the edge of the aperture stop. This way, if the position of a negative lens in an optical system is suitably chosen so that its power is substantial, but the height of the marginal ray on the lens is relatively low, its contribution to the overall optical power is relatively low while still having a significant field curvature.

Two examples where a negative component is effectively used to reduce a field curvature are shown in Fig. 5.23. The first example is the

Figure 5.23
Negatively Powered Elements with Small Value of Y to Flatten Field



Cooke triplet. A negative component is located in the middle between two positive lenses. The marginal ray height on both positive lenses is higher than on the negative lens. However, the power of the negative lens significantly reduces the field curvature created by the two positive lenses.

The second example is a Petzval lens, where the negative component is located very close to the image plane. Its contribution to the power of the whole lens is very small, since the height of the marginal ray is extremely small when the lens is close to the image. If the lens is placed at the image plane, it does not change the overall power of the system.

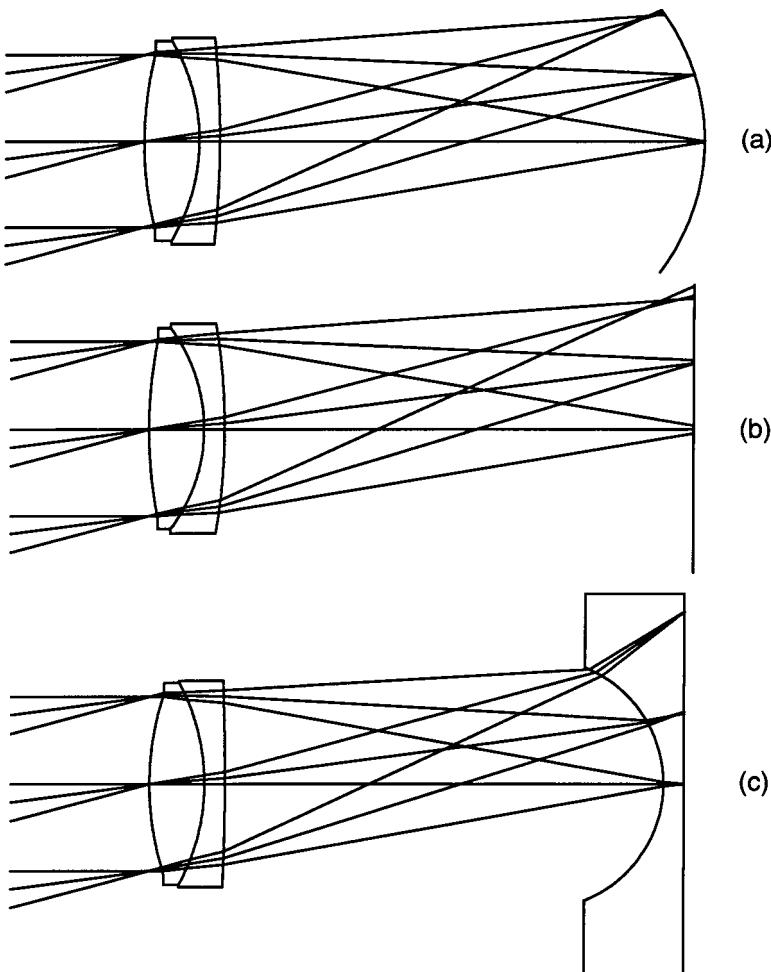
Is there any reason to put the lens at the image plane? Yes, indeed both positive- and negative-powered lenses are often located either in the image plane or very close to it. They are called *field lenses*.

The first case when a lens is located at the image plane or just in front of the image plane is when the negative lens is used as a *field flattener* to correct the field curvature and flatten the field. This is common in complex wide field of view, fast (low $f/ \#$) lenses.

How does a negative lens flatten the field? Let us imagine the case of a simple positive lens that forms an image a certain distance from the lens. If we add a block of glass between the lens and the image plane, the image will move away from the lens. The image shift is proportional to the thickness of the glass block. In the case of the negative lens in front of the image plane, ray bundles that focus close to the edge of the field of view pass through the part of the lens where the glass thickness is larger than in the center of the lens. This way image points that are closer to the periphery of the image are shifted away from the focusing lens more than the ones in the center of the field. This results in a flatter image plane.

A doublet lens forms a sharp image on a spherical surface shown in Fig. 5.24a. If an achromatic doublet is an objective lens of a telescope, which focuses the image on a reticle, or on a CCD detector, it is desirable to correct the field curvature of the lens. The reticle is usually engraved on a flat piece of glass, and in the case of a CCD detector, the sensitive area is always flat. When the CCD detector is adjusted for best focus, both the center and the edge of the field are slightly blurred, as in Fig. 5.24b, and the sharpest image is obtained for the intermediate field. Figure 5.24c shows how a field lens in front of the image reduces the field curvature and the image blur at the edge of the field.

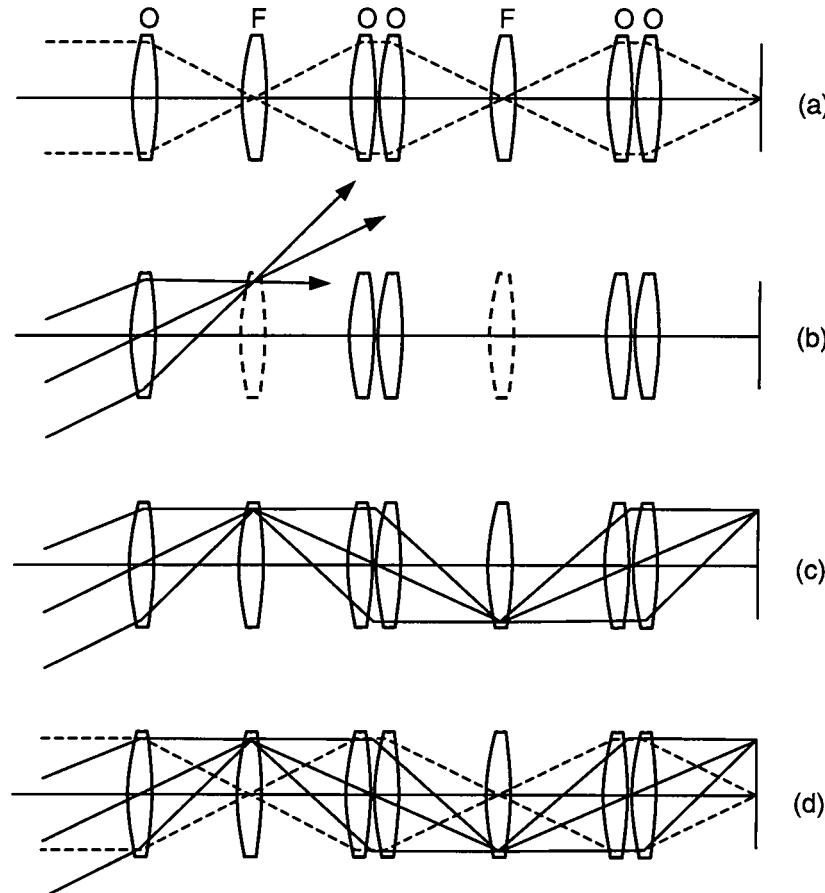
Figure 5.24
Field Flattener Example: How a Negative Lens Flattens the Field



The other types of field lenses are positive-powered lenses used in the systems with one or more intermediate images and relay optics. A submarine periscope is an example of such a system. The optics is inside a tube, which is 10 m long or even longer, but the diameter of the optics does not exceed 250 to 300 mm. Another example is an endoscope, which is on a much smaller scale, but with a similar ratio of the system length and diameter. Schematically, these systems are shown in Fig. 5.25a, where lenses labeled O are objective lenses and lenses labeled F are field lenses.

What is the function of these positive field lenses? They cannot correct the field curvature; they actually *increase* the field curvature already introduced by the other positive lenses. If we look at the axial beam shown in the schematic drawing, and assume from the beginning that

Figure 5.25
Field Lens Example:
How a Positive Lens
Reduces Vignetting

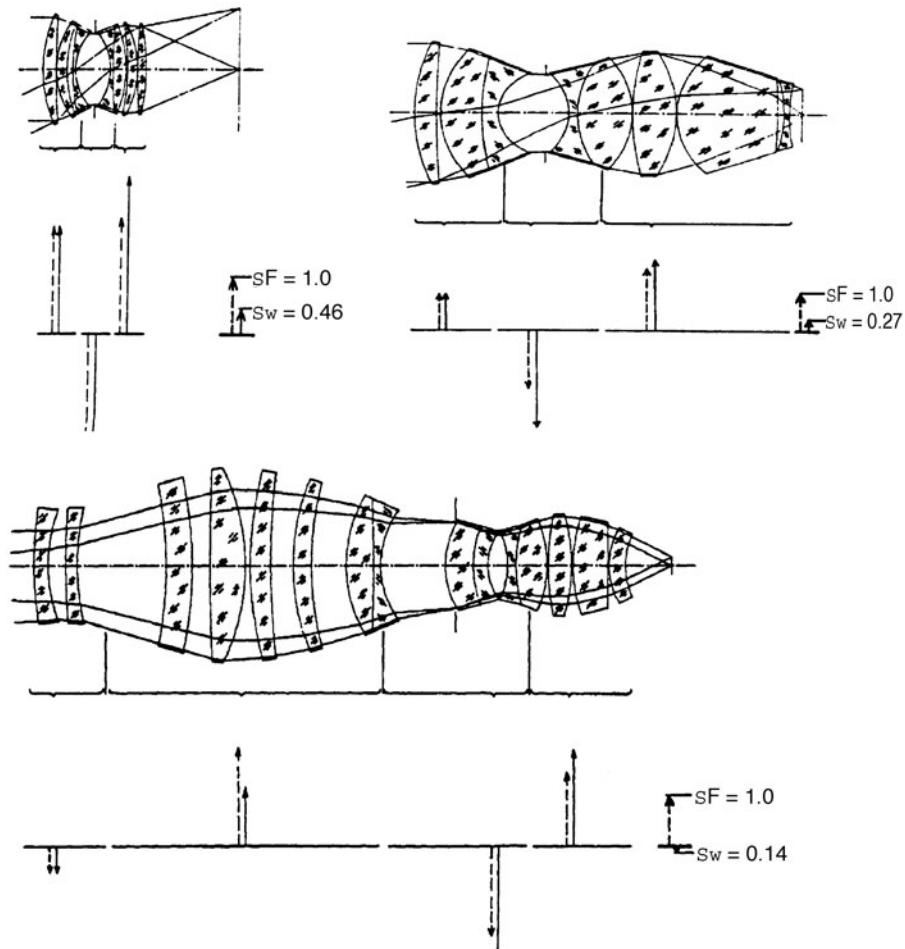


there are no field lenses (only the O lenses are present), the axial beam will be focused at the first intermediate image plane, then relayed with two lenses to the second intermediate plane, and finally relayed with another two lenses to the final image plane. There is no problem with the axial ray bundle. Now consider the beam entering the optical system at an angle. It will be focused at a certain height from the optical axis in the first intermediate image plane. However, the cone of light is so tilted that almost the whole ray bundle is going to miss the two relay lenses, and it will hit the housing. This is called *vignetting* in the optical system and it reduces the amount of light in the image periphery. If we now add a positive lens in the image plane, it will not do anything to the axial beam, but it will redirect the cone of light coming from the edge of the field into the relay lenses. There will be no vignetting and almost no change in the position and the size of the image. The image brightness is going to be uniform across the field, but the system will have a significant amount of field curvature. The primary purpose of these positive field lenses is to reduce or eliminate the vignetting in the system.

There was a paper in 1980 given by Erhard Glatzel of Carl Zeiss in Germany, where he talked about designing lenses in microlithography that imaged a mask onto a 50-mm-diameter silicon wafer at a 5:1 reduction ratio. Microlithography lenses are the most sophisticated lenses in our industry, since they have to resolve submicrometer structures in the flat image, as shown in Fig. 5.26. Glatzel starts from the basic relation for the Petzval sum. Suppose that only one type of glass is used in the entire lens. In order to correct the field curvature, the sum of the powers of all components in the system has to be zero or very close to zero. At the same time, total power of the system has to be positive. The total power is proportional to the sum of weighted powers of the components, where the weighting factor is the normalized height of the marginal ray on the component. Glatzel first analyzes a so-called planar camera lens similar to a double Gauss-type lens. The positive-powered components are the first two and the last two components. The negatively powered components are located in the center of the lens. This power distribution is very similar to a Cooke triplet lens. Below the lens layout, Glatzel plots the contribution to the power sum as well as the weighted power sum from corresponding lens groups above. Petzval curvature in this first lens is not reduced to zero but has a residual of 0.46.

In order to reduce the Petzval sum or field curvature, Glatzel forces the central elements to have a more negative optical power. The Petzval sum is lower than in the first example but it is still present. Unfortunately, the

Figure 5.26
Reduced Field Curvature Lenses as
Described by Glatzel
of Carl Zeiss



radii of the central elements have become quite strong or severe, thus increasing the angles of incidence on their surfaces (especially the short-radius inner concave surfaces), therefore introducing their own aberrations.

The final step in the design of the flat-field lithography lens is the addition of a negative group in the front of the whole lens. This group has a very important role. Because of its negative power, it slightly reduces field curvature, but it expands the beam and makes it possible for the following positive group to make a high contribution to the total power of the system. The final lens design has -0.14 of field curvature.

While this still has a noticeable result, the net aberration of the system is extremely well corrected. This form of lens still represents the state of the art in lithography optics.

While we are discussing this design, we should take a few minutes to look at the design form. After the initial two negative elements in the front, Glatzel uses four larger single elements to take the diverging light and converge it into the next group. He uses four elements in order to split the power and minimize angles of incidence so as to reduce aberrations. However, note that the second element seems to have most of the positive power; also the first element of this group seems to be quite concentric about the diverging light cone, and it seems to have little or no optical power. We point this out in order to suggest that this element may indeed be able to be removed from the design. This is something that you should always be looking for during the optimization of your design in order to, where possible, simplify the design. It is difficult, if not impossible, to make a definitive judgment on this without working with the specific design data and optimization. It is certainly possible that this element may be canceling some higher-order aberration residual.

Figure 5.27 shows two photos of a 17-element lithographic lens produced by Corning/Tropel. It is clear how the individual lens elements are bent in order to minimize the ray angles of incidence on many of the surfaces which, in turn, will tend to reduce the residual aberrations. This is especially clear in the 30-element lens in the lower figure. These lenses are typically extremely difficult to manufacture, assemble, and align due to the large number of surfaces.

Distortion

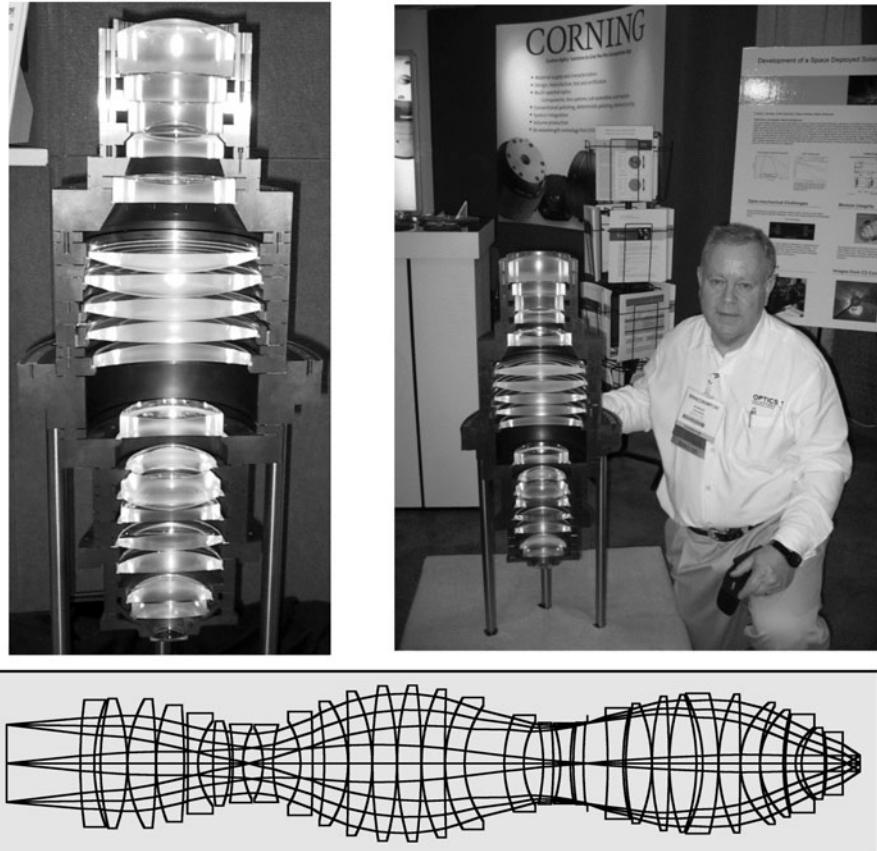
The only aberration that does not result in image blur is distortion. If all other aberrations in the system, except distortion, are corrected, an object point is imaged onto a perfect image point, which is displaced from its paraxial position. The amount of distortion can be expressed either as a lateral displacement in length units or as a percentage of the paraxial image height. Distortion is defined as

$$\text{Distortion} = \frac{y - y_p}{y_p}$$

where y is the height in the image plane and y_p is the paraxial height.

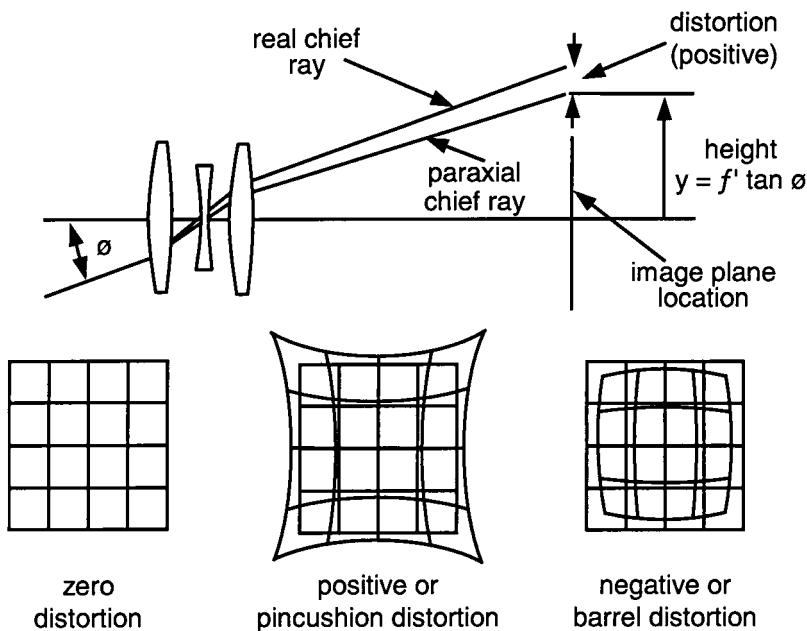
Figure 5.27

A 30-Element Lithography Lens from the Patent Literature



Third-order distortion increases with the cube of the field of view. A distorted image of a rectilinear object is shown in Fig. 5.28. Distortion can be *positive or pincushion distortion* or alternately *negative or barrel distortion*. For a thin lens with the aperture stop on the lens, distortion is equal to zero. The thickness of a lens and its position relative to the aperture stop determines its contribution to the system distortion. An example of a system where a correction of the distortion is a big challenge is a wide field-of-view eyepiece. Its field of view may be as high as 70° , and its aperture stop, which is the pupil of the eye, can be in the order of 20 mm away from the system. If the object and the image are interchanged, the lens that has a barrel distortion in one direction has a pincushion distortion in the opposite direction.

Figure 5.28
Distortion



This is a very interesting subtlety, and we urge the reader to spend due time thinking about it prior to predicting what a distorted image will look like to make sure your sign of distortion is modeling the real world properly.

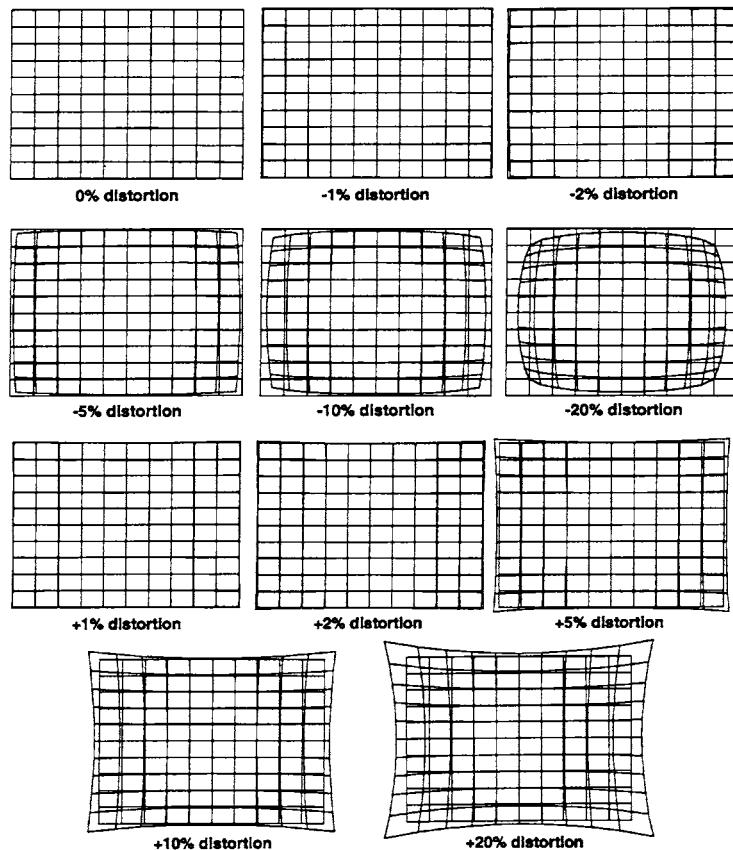
Figure 5.29 shows different amounts of negative and positive distortion. Since distortion is a cosmetic-type aberration not affecting resolution, its appearance is very important, especially in visual systems. Generally distortion in the order of 2 to 3% is acceptable visually.

Figure 5.30 shows where distortion comes from. In this situation, the aperture stop is located to the left of the lens, and the angle of incidence on the lens by the ray bundle is large enough so that there is a reasonable difference between the paraxial angle of refraction and the real ray angle of refraction. As with spherical aberration, the real rays are refracted more severely than the paraxial rays. In this case, this causes the real image to be pulled inward from the paraxial image thus causing negative or barrel distortion.

Table 5.1 summarizes the aperture and field dependence of the primary aberrations.

Figure 5.29

Illustration of Different Amounts of Negative and Positive Distortion

**Figure 5.30**

Where Does Distortion Come From?

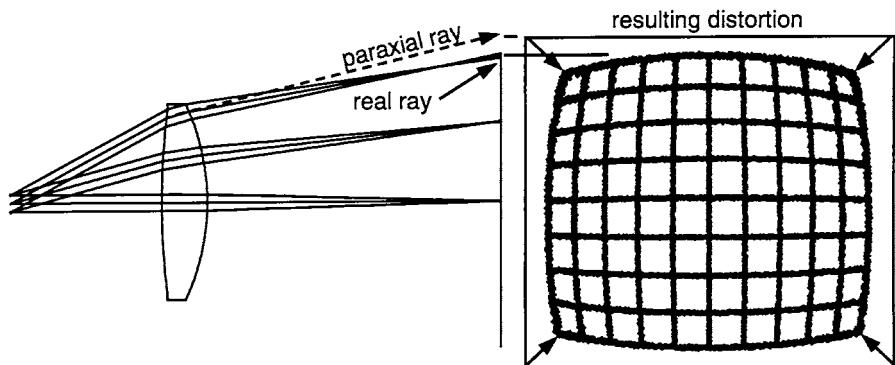


TABLE 5.1

Summary of Third-Order Monochromatic Aberration Dependence on Aperture and Field Angle

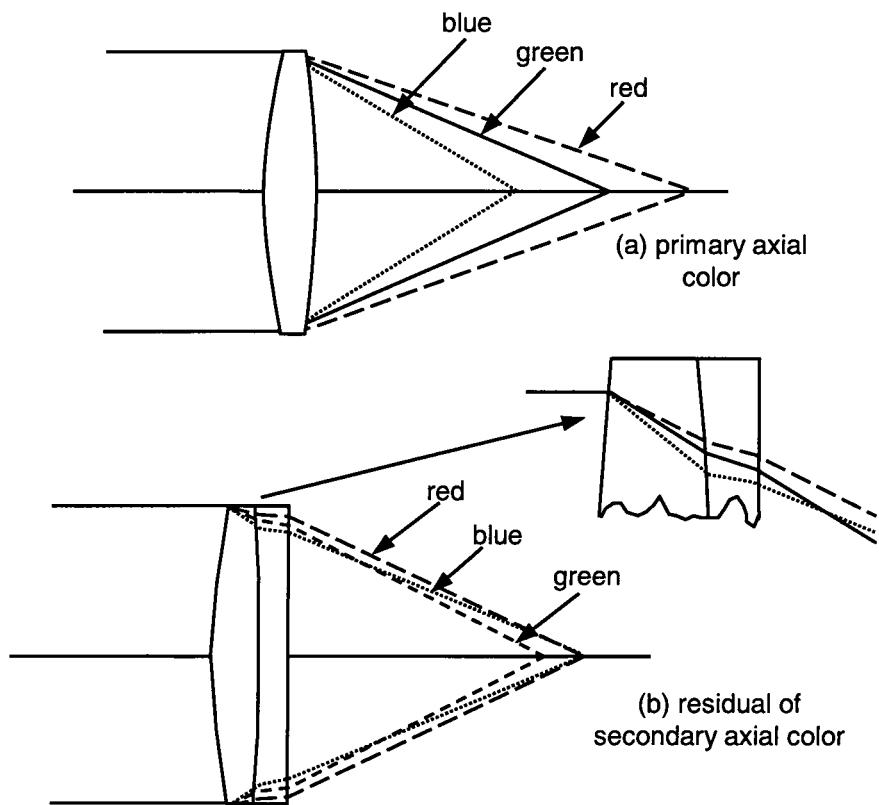
	Aberration	Aperture Dependence	Field Dependence
	Spherical	Cubic	—
	Coma	Quadratic	Linear
	Astigmatism	Linear	Quadratic
	Field curvature	Linear	Quadratic
	Distortion	—	Cubic

Axial Color

If white light is incident onto a glass wedge or a prism, it is decomposed into a rainbow. This is called *dispersion*. Blue light is refracted more severely than red light, since the index of refraction is higher for shorter wavelengths than for the longer wavelengths. The properties of lenses also vary with wavelength. White light coming from an axial infinitely distant object, which is incident upon a convergent lens, is shown in Fig. 5.31a. The edge of the lens acts like a wedge, refracting or bending the blue light more than the red light. This causes the blue light to focus closer to the lens than the red light. This longitudinal variation of focus with wavelength is called the *axial chromatic aberration* or *axial color*. In the absence of spherical aberration, a system with uncorrected chromatic aberration forms a bright spot surrounded with a purple halo coming from the blue and red light.

Is there a way to correct the axial color? A lens that focuses an infinitely distant object is shown as an example in Fig. 5.31b. In order to bring the blue and the red to focus together, a positive lens must be split into two lenses made of glasses with different dispersions. The first is a positive lens with low dispersion glass. This type of glass is called a *crown* glass. The second lens has a lower optical power than the first one, so that the total power of the doublet is positive. However, the second lens is made of a high-dispersion glass called a *flint* glass, which means that it spreads light more with color, and it cancels most of the axial chromatic aberration created by the first lens because of its negative power. This doublet is called an *achromatic doublet*.

Figure 5.31
Axial Chromatic
Aberration



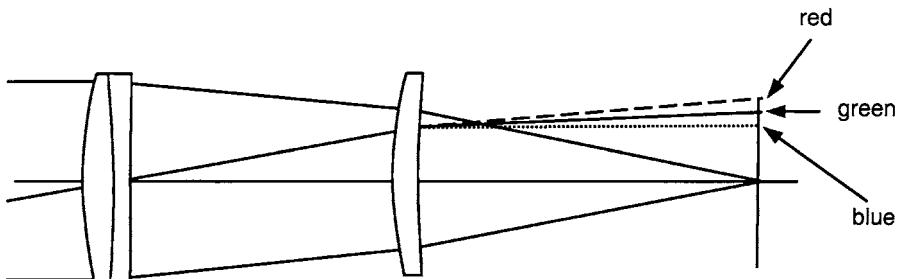
Lateral Color

When a lens forms an image of an off-axis point at different heights for different wavelengths, the lens has lateral chromatic aberration or chromatic difference of magnification. This aberration is quite common in wide field-of-view systems. A very descriptive name for lateral color is *color fringing* since this is what is seen when looking at an image formed by a lens with lateral color.

Lenses that are further from the aperture stop in a system contribute more to lateral color than the lenses with smaller chief ray heights. Lateral color created by a lens is shown in Fig. 5.32. The chief ray is going through the single-lens element close to its outer periphery. Shorter wavelengths are bent or refracted more severely than the longer wavelengths. The blue image is formed closer to the optical axis than the red image. In



Figure 5.32
Off-Axis Lateral Color
(Color Fringing)



wide field-of-view systems, lateral color is often the aberration that is the most difficult to correct. Its correction may require the use of anomalous dispersion glasses, which are often expensive, or diffractive elements.

Visual systems often have lateral color due primarily to the eyepiece, which inherently has a lot of lateral color. If you look, for example, through a pair of binoculars at a sharp bright/dark edge close to and tangent to the edge of the field of view, you will likely see severe lateral color. This is often not a problem, as the user will most often place the object of interest at the center of the field of view. For this reason visual optical systems tend to be somewhat more forgiving than other systems.

Parametric Analysis of Aberrations Introduced by Plane Parallel Plates

Let us assume that we have a diffraction-limited $f/1$ lens. What is the spherical aberration introduced by a plane parallel plate inserted in the converging cone between the lens and the image plane? The spherical aberration introduced by the plate increases as a function of the plate thickness. Figure 5.33 shows how the blur diameter due to the spherical aberration changes as a function of the $f/\#$ of the lens and the plate thickness. The shaded area is the region where the optical system is diffraction limited. For our $f/1$ lens, the Airy disk diameter is $1.5 \mu\text{m}$. If a plane parallel plate of glass 50 mm thick is inserted in the $f/1$ cone of light, the spherical aberration in the focused image increases to 1.25 mm . The image blur is much larger than the Airy disk, which means that the glass plate significantly degrades the performance. We would have to stop down the aperture of the lens to about $f/5$ to reduce the size of the

Figure 5.33
Spherical Aberration
of a Plane Glass Plate,
Refractive Index =
1.517 (BK7)

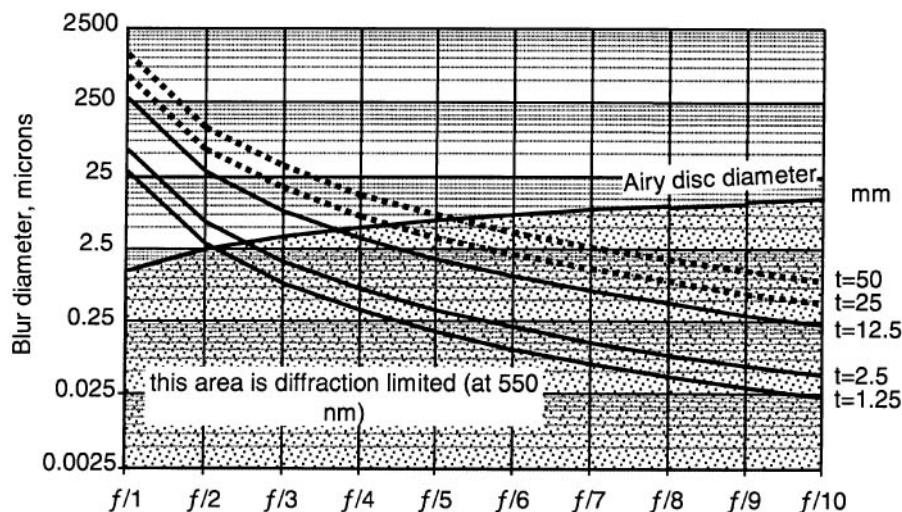


image blur to the Airy disk. At the point where the diffraction blur and the aberrations of the glass plate are equal, the glass plate will not have a detrimental effect on the image quality.

Third-order spherical aberration, coma, astigmatism, and distortion of a plane parallel plate depend on the index of refraction of the plate, n , and the thickness, d , and they are proportional to

$$\frac{(n^2 - 1) d}{n^3}$$

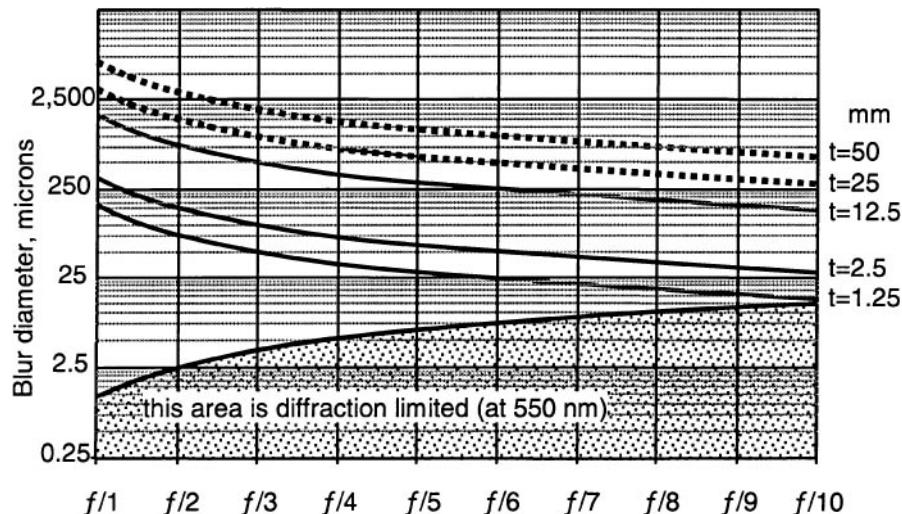
What happens if a plane parallel plate is inserted in a converging cone of light at 45° ? If the tilted plate is inserted in a rotationally symmetric optical system in a convergent beam, the optical system is no longer rotationally symmetrical. A result of this is the presence of field aberrations such as coma, astigmatism, and lateral color in the on-axis field position.

Let us again assume that we have a diffraction-limited, very fast $f/1$ lens and we need to split the beam into two beams, one reflected at 90° and the other beam transmitted through the tilted plate beamsplitter. The reflected beam is unchanged after reflection off the plate, but the transmitted beam has on-axis astigmatism introduced by the tilted plate. The thicker the plate, the more astigmatism is present. The blur diameter associated with third-order astigmatism as a function of the thickness of the tilted plate and the $f/ \#$ of the lens is shown in Fig. 5.34.

Is there a way to correct this on-axis astigmatism? There are a few different viable methods of correction. Astigmatism is proportional to the

Figure 5.34

Third-Order Astigmatism Blur Diameter, in Micron, as a Function of the Thickness of a 45° Tilted BK7 Plate and the $f/ \#$ of the Lens



square of the tilt angle, so it is therefore not dependent on the sign of the plate tilt. Thus, astigmatism cannot be compensated with a second plate tilted in the opposite direction from the first plate and in the same plane of tilt. This is shown in Fig. 5.35. However, if the second plate is tilted in a plane which is orthogonal to the plane of tilt of the first plate, the astigmatism can be corrected for the most part. The second method of correction involves the use of a weak spherical surface on the tilted plate or a weak wedge instead of the plane parallel plate. Although it is more difficult for fabrication, a good correction of the astigmatism can be achieved with a decentered cylindrical surface on the tilted plate. You can also think of this component as a wedged cylinder. In this way, the astigmatism, as well as a smaller residual of coma, are reasonably well corrected.

The most severe aberration introduced in an optical system by a tilted plate is astigmatism. However, coma and lateral color are also significant in the case of a fast lens even when the plate thickness is less than 1 mm. The

Figure 5.35

Correction of Astigmatism from a Tilted Plate in a Converging Beam

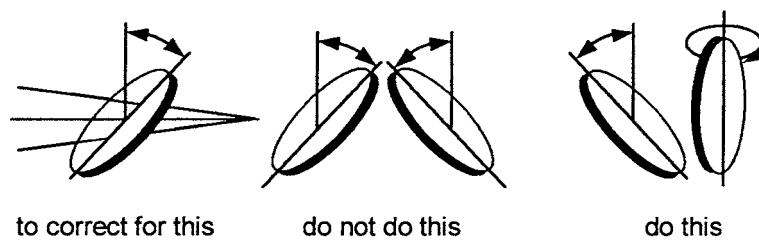
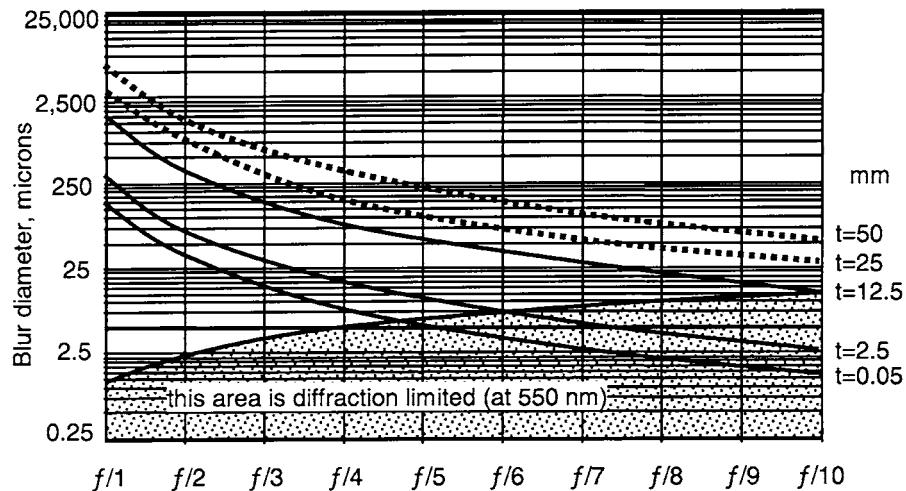


Figure 5.36

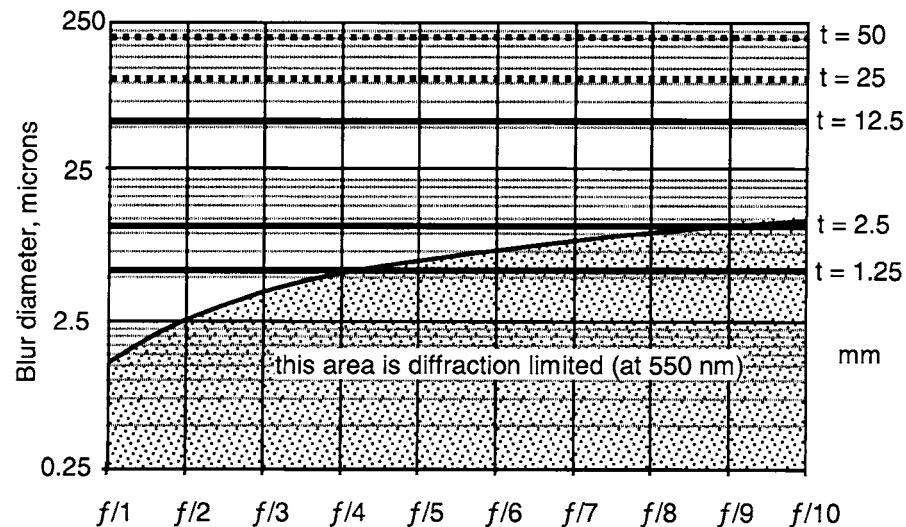
Tangential Coma Blur Diameter As a Function of the Thickness of a 45° Tilted Plate of Index 1.5 and the f/# of the Lens



residual tangential coma blur in a system with a 45° tilted plate is shown as a function of plate thickness and the f/# of the lens in Fig. 5.36. Lateral color blur in a system with a 45° tilted plate is shown as a function of plate thickness and the f/# of the lens in Fig. 5.37. Note that the lateral color is independent of the f/number. This is because the lateral color is only dependent on the height of the chief rays in the different wavelengths.

Figure 5.37

Lateral Color Blur Diameter in Microns as a Function of the Thickness of a 45° Tilted BK7 Plate and the f/# of the Lens



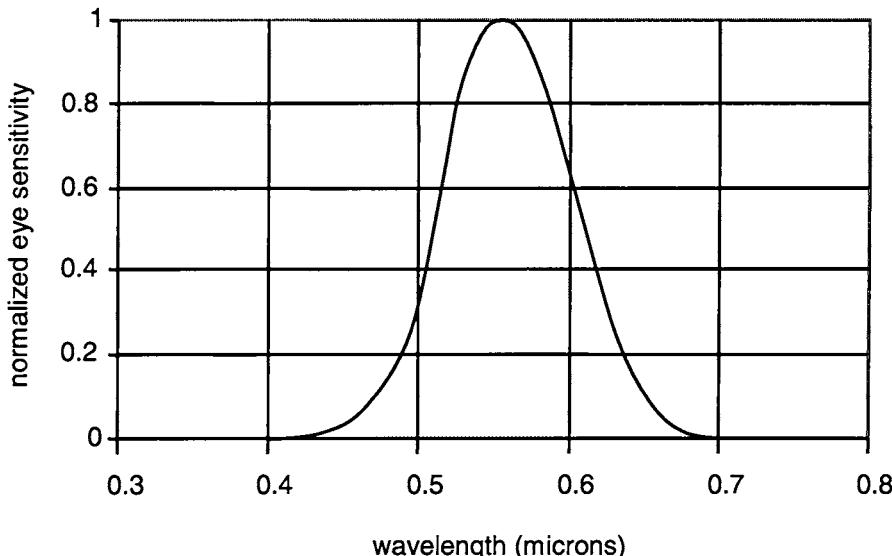
Glass Selection (Including Plastics)

Material Properties Overview

Every optical system works in its own particular wavelength region determined by the spectral characteristics of the light source, the spectral sensitivity of the sensor, as well as any other factors or components which alter the net sensitivity of the system. If an optical system is a visual system, the optical materials must be transmissive between approximately 425 and 675 nm, as determined by the photopic spectral response curve of the human eye. The photopic eye sensitivity is shown in Fig. 6.1. Optical glasses are the most commonly used materials in optical systems. However, there are some optical plastics with good transmission in the visible spectrum that can be injection molded. In high-volume production, this technology is significantly cheaper than classical glass manufacturing methods. Operating temperature range is very important when choosing optical materials. Optical materials change their index of refraction with temperature, and they also expand differently, changing the lens shape and optical power. Optical plastics have approximately one order of magnitude higher coefficient of thermal expansion than glasses.

If the temperature in an optical system rises to a few hundred degrees Celsius, plastic materials cannot be used because the plastic will melt. Most optical glasses can withstand temperatures of a few hundred degrees Celsius without changing their shape. In illumination systems close to the light source, the temperature can go up to 900°C. In this

Figure 6.1
Photopic Spectral Eye
Sensitivity Curve



case, glass optics will melt too. Fused quartz or fused silica is often used in these systems because it can operate at temperatures close to 1000°C.

Manufacturers of optical components usually include in their catalogues information about the standard optical materials they use. An example of the general information on optical materials for the visible, near ultraviolet (UV), and near infrared (IR) spectral regions as provided by Melles-Griot is shown in Fig. 6.2.

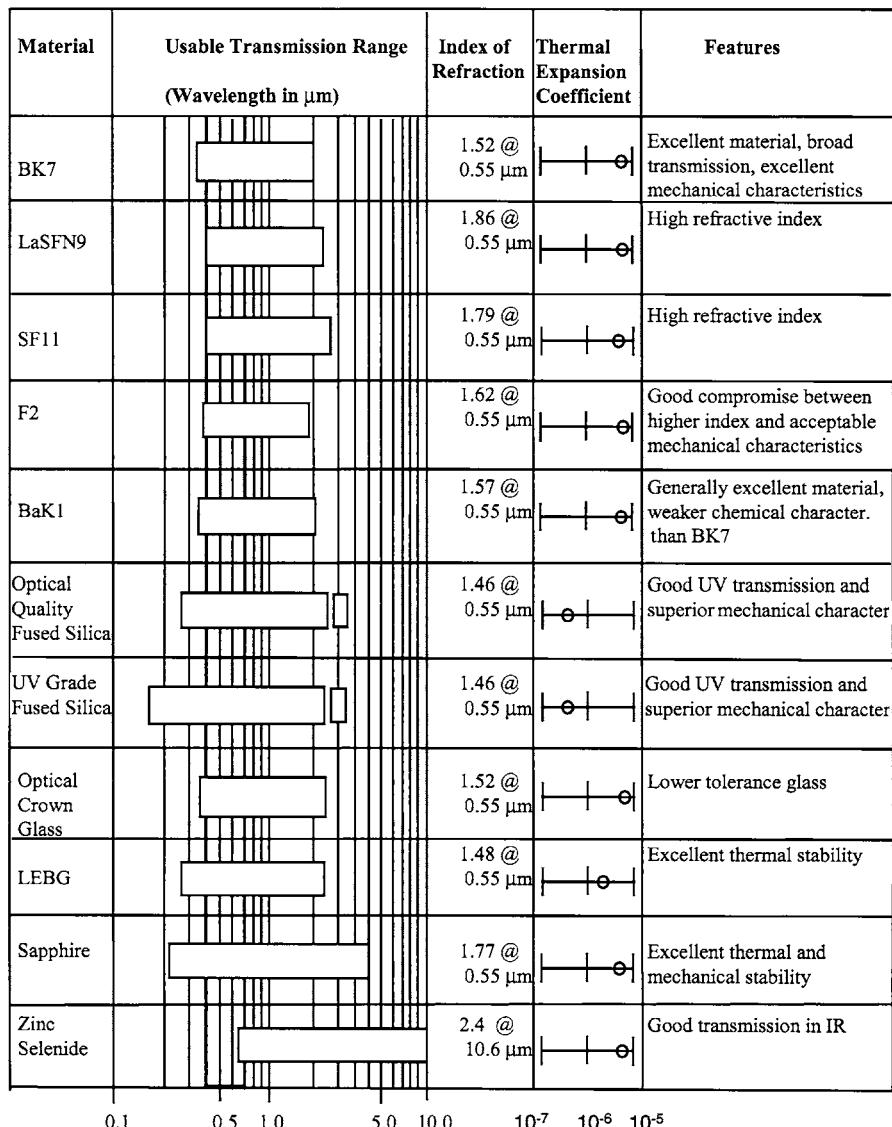
The Glass Map and Partial Dispersion

The refractive index of all optical materials changes as a function of wavelength. The refractive index increases as the wavelength decreases. This means that optical systems with refractive components have chromatic aberrations. In fact, the performance of an optical system is often limited by chromatic aberrations rather than monochromatic aberrations.

In the time of Sir Isaac Newton, it was believed that it was not possible to correct chromatic aberrations by combining different types of glasses. Newton thought that the chromatic aberrations of all lenses were

Figure 6.2

Material Properties Overview from
Melles-Griot Catalog
(Simplified)



proportional to their powers, with the same constant of proportionality for all glasses. This is the reason why Newton built a reflecting telescope. In the eighteenth century, however, it was found that, with the proper choice of glasses and powers, it was possible to design an *achromatic doublet*, which was chromatically corrected for two wavelengths.

Let us consider two thin lenses made from two different glass types and cemented together as shown in Fig. 6.3. We want to find the condition for this doublet to be an achromatic doublet, chromatically corrected for the red C line wavelength 656.27 nm and for the blue F line 486.13 nm. Generally, the *crown* materials are less dispersive than the flints, and in an achromatic doublet we combine the less dispersive crown as the positive element and the more highly dispersive *flint* as the negative element. The central wavelength is usually chosen as the *d* line, which is 587.56 nm. If the power of the first lens is P_1 and the second lens P_2 , then the total power of the doublet is

$$P = P_1 + P_2$$

An achromatic doublet will have the same power for the *C* line wavelength and the *F* line wavelength if

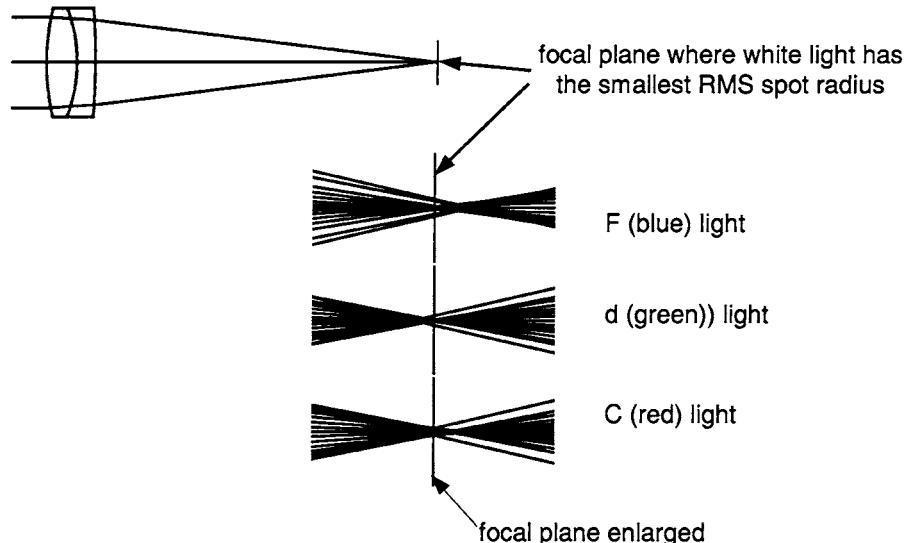
$$(P_1 + P_2)_C = (P_1 + P_2)_F$$

or

$$P_1 \frac{n_{1C} - n_{1F}}{n_{1d} - 1} = -P_2 \frac{n_{2C} - n_{2F}}{n_{2d} - 1}$$

Figure 6.3

Focusing of White Light with an Achromatic Doublet from BK7 and SF2 Glasses



where P_1 and P_2 are the powers of two thin lenses at 587.56 nm. The value $(n_F - n_C)$ is called the *main dispersion*. The ratio

$$V = \frac{n_d - 1}{n_F - n_C}$$

is called the V number or the Abbe number. The condition for a doublet to be an achromatic doublet becomes

$$\frac{P_1}{V_1} = -\frac{P_2}{V_2}$$

From this relation we can obtain the focal lengths of two components of an achromatic doublet as

$$f_1 = f \frac{V_1 - V_2}{V_1}$$

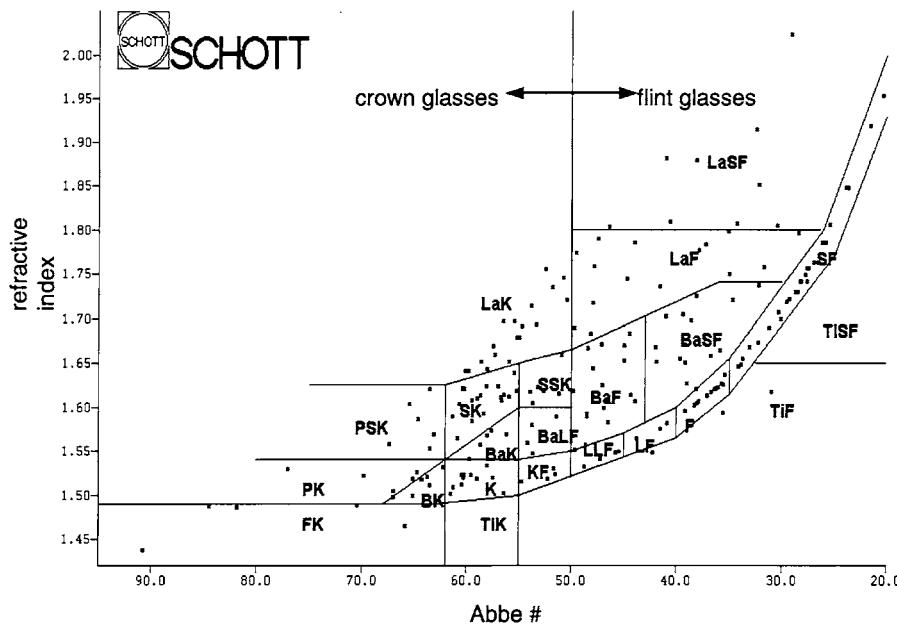
$$f_2 = -f \frac{V_1 - V_2}{V_2}$$

where f is the focal length of the doublet. The net result of this is to derive the powers of the less dispersive crown element and the more dispersive flint element, so that the combined doublet focal length in the red and blue wavelengths are the same. When this condition is reached, the central wavelength (green) is defocused slightly toward the lens.

In the second half of the eighteenth century, Ernst Abbe worked closely with Otto Schott on testing different types of optical glass and this encouraged the development of new glass types. It was found that the most suitable way to characterize optical glass was the specification of the index of refraction for the d line, n_d , and the Abbe number, V , which determines the glass dispersion. Manufacturers of optical glasses provide a glass map or an n_d/V_d diagram in which the Abbe number is plotted as the abscissa and the index, n_d , as the ordinate. The glass map from the Schott glass catalogue is shown in Fig. 6.4. Schott is the largest manufacturer of glass in the world, but there are other manufacturers, including Hoya, Ohara, Pilkington, Corning, and Sovirel.

The n_d/V_d diagram subdivides the various types of glasses into groups, each having a specific designation such as BK with BK7, BK1, and others. These designations are generally related to the fundamental materials used in the manufacture of the specific group such as LAFN31 which is a lanthanum flint glass. There is also a more general division of glasses into "crown" and "flint" glasses. The crown glasses are

Figure 6.4
Schott Glass Map
(Abbe Diagram)



the ones with $n_d > 1.60$, $V_d > 50$ or $n_d < 1.60$, $V_d > 55$; the other glasses are flints. The available refractive indices range from 1.45 to 2 and the V number from 80 to 20.

Mathematically, the dependence of refractive index on the wavelength of light can be expressed in a few different ways, but none of the expressions is highly accurate over the entire glass transmission range. These relationships are empirically derived from measured data. The Sellmeier dispersion formula is

$$n^2 - 1 = \sum_i \frac{1}{\lambda^2 - \lambda_i^2} c_i \lambda^2$$

and the formula from the 1967 Schott catalogue is

$$n^2 = A_0 + A_1 \lambda^2 + \frac{A_2}{\lambda^2} + \frac{A_3}{\lambda^4} + \frac{A_4}{\lambda^6} + \frac{A_5}{\lambda^8}$$

These are the two most commonly used formulas. Besides these two formulas, the Hartmann and Conrady formulas are also offered in some of the lens design programs. The six constants that characterize glass dispersion vary considerably between glasses, and thus the general shapes of all dispersion curves are different.

In addition to the main dispersion ($n_F - n_C$), which is the difference between the index of refraction for the blue and for the red line, the “partial dispersion” is also commonly used. Partial dispersion in the blue is the difference in index of refraction between 435.83 and 486.13 nm, and the red partial dispersion is the difference in index of refraction between 653.27 and 852.11 nm.

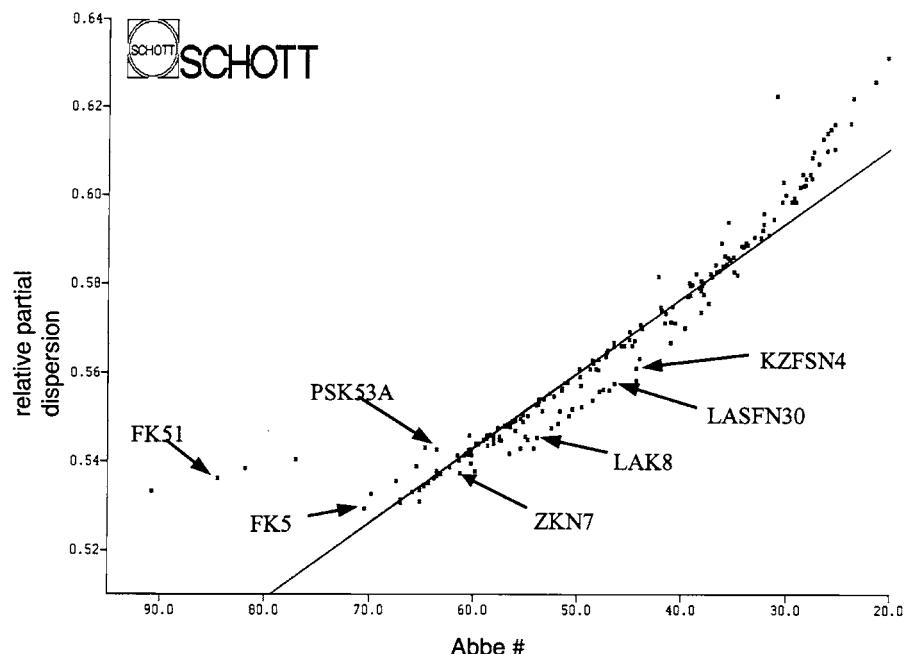
Perhaps even more important is the “relative partial dispersion,” which is the ratio of the partial dispersion and the main dispersion. Generally, the relative partial dispersion is

$$P_{x,y} = \frac{n_x - n_y}{n_F - n_C}$$

A glass map with the relative partial dispersion as a function of the Abbe number from the Schott catalogue is given in Fig. 6.5.

The derived formulas for the design of an achromatic doublet provide a chromatic correction for *F* and *C* wavelengths. However, dependent on the choice of glasses, there will be a residual mismatch of dispersions, resulting in a larger or smaller “secondary spectrum.” This secondary spectrum is the difference in image position between the central wavelength (green

Figure 6.5
Relative Partial Dispersion



or yellow) and the now common blue and red image position. In order to eliminate the secondary spectrum, we should find a pair of glasses with different V values but the same relative partial dispersion. Abbe showed that the majority of glasses, the so-called normal glasses, exhibit an approximately linear relationship between the relative partial dispersion and the Abbe number. Thus

$$P_{x,y} \approx a_{x,y} + b_{x,y} V_d = (P_{x,y})_{\text{normal}}$$

which can be clearly seen in the relative partial dispersion map in Fig. 6.5. The reduction of the secondary spectrum requires the use of at least one glass type which does not lie on the $(P_{x,y})_{\text{normal}}$ line. The glasses that lie away from the line of normal glasses are often expensive and may be difficult to manufacture. Some of them are shown in the relative dispersion chart in Fig. 6.5. KZFSN4 is seven times more expensive than BK7, LaK8 is eight times, PSK53A is 11 times, and LaSFN30 24 times.

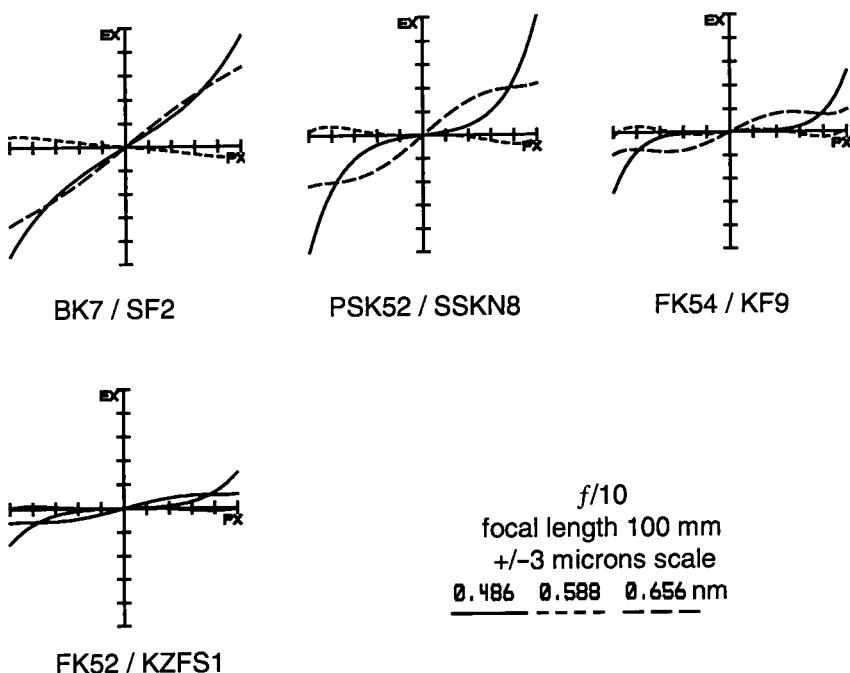
Parametric Examples of Glass Selection

In this section we will show how secondary spectrum can be corrected and to which level, with the right choice of glasses. It will also be shown how the spherical aberration and the secondary spectrum of a doublet are dependent on the $f/\#$. These data are shown as parametric analyses.

The first parametric study is shown for the case of an $f/10$ achromatic doublet using different glass combinations. Four doublets will be compared, the first using two normal glasses, and then using anomalous dispersion glasses for one or both of the elements. The first doublet is designed with two normal glasses: BK7 and SF2. It is an $f/10$ lens with a 100-mm focal length. The ray aberration curves are shown in Fig. 6.6. An explanation of these curves will be given in Chap. 10. The difference in the Abbe numbers between the two glasses should be sufficiently large, so that the shape or the power of each individual component is reasonable. Note that as the Abbe number difference between the two glasses of an achromatic doublet decreases, the relative powers of the positive and negative elements get stronger. This yields greater spherical aberration, as will be seen. For the $f/10$ lens the spherical aberration for the central wavelength is very small, the rms spot diameter is less than 1 μm . However,



Figure 6.6
Secondary Spectrum
Correction As a
Function of Glass
Selection



both the red and the blue foci are away from the green focus, which means that the secondary spectrum aberration is not corrected.

The second achromatic doublet is designed with PSK52 and SSKN8 glasses. SSKN8 is a normal glass, but PSK52 has anomalous dispersion. The green spot diameter is the same as in the case of the first doublet, but the polychromatic spot diameter is slightly smaller because the secondary spectrum is lower. Note here that we have an increase of spherical aberration which changes with wavelength. The blue has positive spherical aberration and the red has negative spherical aberration. This change in spherical aberration with wavelength is called *spherochromatism*.

The third case is a doublet with FK54 and KF9 glasses. Although KF9 is a normal glass, FK54 has an extremely high anomalous dispersion, resulting in a much better secondary spectrum correction than in the previous case. FK54 is over 30 times the cost of standard BK7 glass!

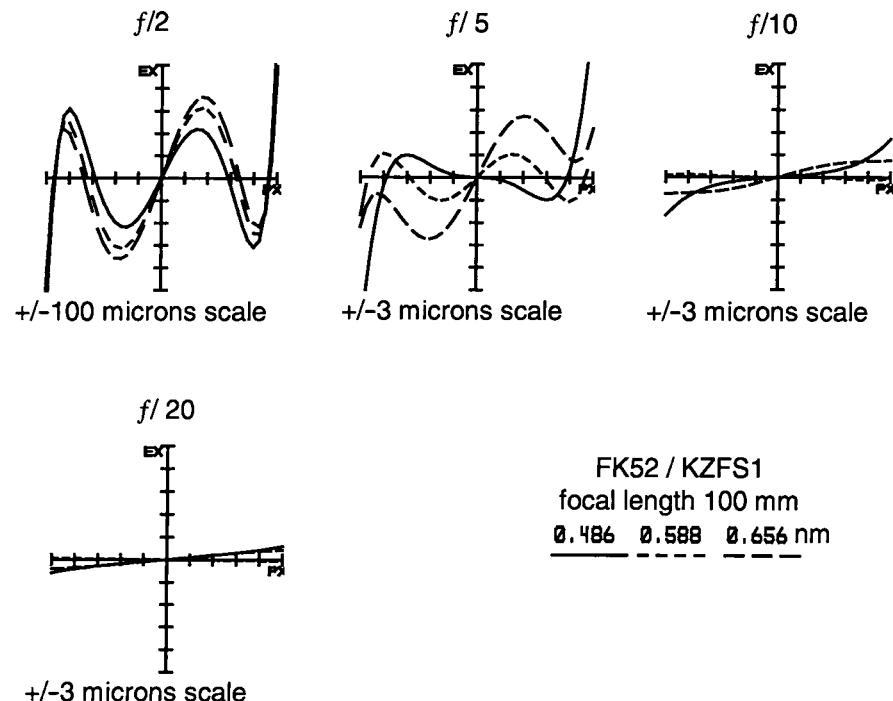
The fourth case is a doublet with FK52 and KZFSN4. Both glasses have anomalous dispersions. The polychromatic spot diameter is less than 1 μm , and the secondary spectrum is completely corrected. Furthermore, the spherochromatism is extremely well corrected as well.

The second parametric study is the analysis of spherical aberration and secondary spectrum correction as a function of the lens $f/ \#$ for a chosen set of glasses.

The first pair of glasses are FK52 and KZFS1, as shown in Fig. 6.7. As demonstrated for an $f/10$ lens, the secondary spectrum can be very well corrected, since both glasses have anomalous dispersion. The Abbe numbers for these two glasses are very different; however, the relative partial dispersion is not. In the case of the $f/2$ lens, the spherical aberration is dominant, with the spot diameter close to 200 μm . As the $f/ \#$ increases to $f/5$, the spot diameter decreases dramatically to about 3 μm . Chromatic aberration is more pronounced in this case. As the $f/ \#$ increases to $f/20$, both the spherical and the chromatic aberrations are extremely well corrected, and the spot diameter is less than 1 μm .

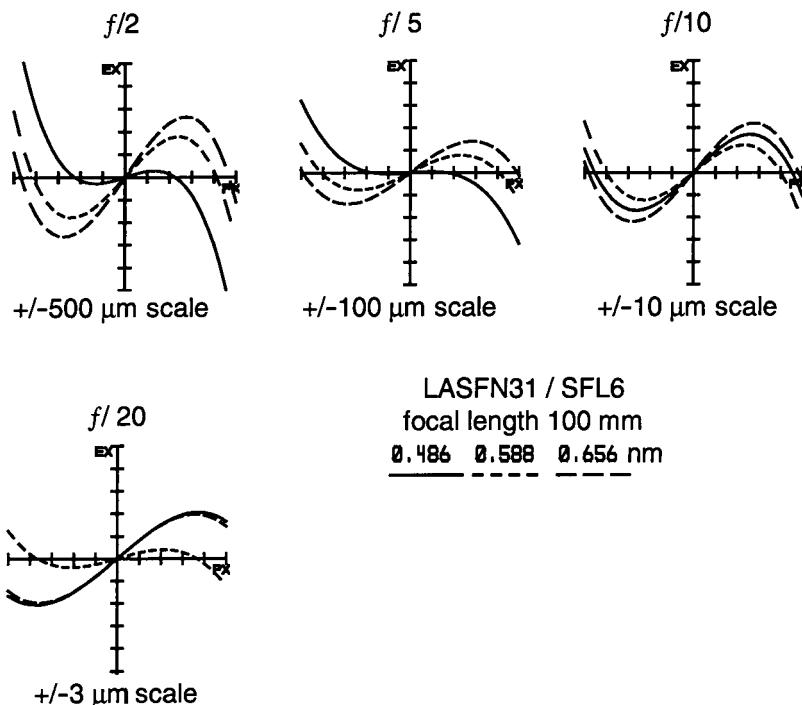
The second pair of glasses to be considered is LASFN31 and SFL6, as shown in Fig. 6.8. Although these glasses are high-index glasses and have anomalous dispersion, their relative partial dispersion is quite different and the secondary spectrum correction is poor. The difference in the

Figure 6.7
Spherical Aberration
and Secondary Spec-
trum Correction As a
Function of $f/ \#$



**Figure 6.8**

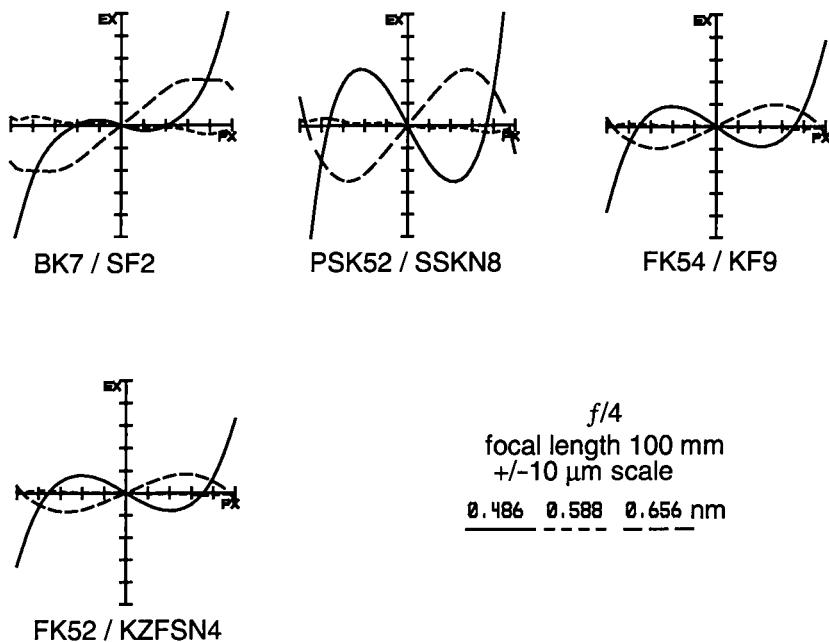
Spherical Aberration and Secondary Spectrum Correction As a Function of $f/\#$



Abbe numbers is not large. This makes it difficult to correct for the spherical aberration too, since we need to correct both aberrations simultaneously, trying to find the optimum balance between the two. In the case of the $f/2$ lens, the spot diameter is around 400 μm . As the $f/\#$ increases, the spot diameter decreases to about 2 μm at $f/20$, largely because of the uncorrected secondary spectrum.

The third parametric study is done for the case of an $f/4$ achromatic doublet using different glass combinations but allowing one surface to be aspheric for nearly complete correction of spherical aberration. This allows us to better see the change in residual spherical aberration with wavelength or spherochromatism. Four doublets are compared, with the same glass combination as the first parametric study. Ray aberration curves are shown in Fig. 6.9. In all four cases, the spherical aberration for the central wavelength is almost perfectly corrected with the asphere on the front surface of the doublet. The spot diameter is determined by the chromatic blur of the blue and the red wavelength, and the difference between four cases of glass combination is small. Note that the shape of the ray aberration curves for the blue and the red color are similar in all cases.

Figure 6.9
Secondary Spectrum Correction As a Function of Glass Selection with One Aspheric Surface



How to Select Glass

Let us imagine that we have to design a wide field-of-view objective that operates in the visible spectrum. We will most likely start with five to seven elements and select a starting configuration based on our prior work, a patent, or we may elect to derive the design from basic principles. In most cases, it is sufficient to optimize a lens for only three wavelengths if the system is a visual system. However, in the case of a large field of view, it is better to work with five wavelengths properly weighted and a larger number of field points because of the potential problems with lateral color correction. At some point in the design, we will start to change the glass types either manually or by varying the glass characteristics automatically in the optical design program, and allowing them to move across the glass map until they settle in the locations which provide the lowest merit function. We will notice that the glasses tend to go from FK, PSK, across SK, LaK, LaSF, to SF. Very rarely a chosen glass will be KE, LLF, LF, or F. Even the glasses, such as BaLF, BaE, or BaSF are not so often chosen in the optimization process. However,

there are some types of glasses that are in the central region of the glass map such as KzFS and TiF that are often chosen in the glass optimization process. The reason for this is that, unlike BaF or BaLF glasses, which are the normal glasses, these glasses have anomalous relative partial dispersions and color correction, particularly secondary spectrum, is much easier with these glasses.

Generally, glasses in the central region of the glass map (BaLF, BaE, BaSF glasses) are not frequently used because they are normal glasses and secondary spectrum is not easily corrected with them. The other reason is that their V number has a medium value, which means that primary color is not easily corrected either. The exception is KzFS glasses, which have anomalous dispersion. Unfortunately, as will be discussed later, KzFS glasses are not preferred Schott glasses.

Now comes a very important step in the optical design—we have to check many parameters for each glass, including its availability, price, transmission properties, thermal properties, staining, etc., and make sure that the glass choice is the optimum one. Here the most important parameters an optical designer has to consider in the process of glass selection will be described. This will also be covered in Chap. 18 from a manufacturer's perspective.

AVAILABILITY Glasses are divided into three groups: preferred, standard, and inquiry glasses. Preferred glasses are always available. Note that just because a glass is "preferred" does not mean that it is of good optical characteristics or low in cost, nor does it mean that the glass is easy to work in the shop; it only refers to availability. Quick delivery of standard glasses is generally possible as these glasses are generally in inventory. Inquiry glasses are available only on request, and they are normally not in stock. An optical designer should make every effort to design the system with preferred glasses. The optical design software program Zemax contains an option to use only preferred glasses from the Schott catalogue when a system is optimized with the "hammer" optimization and "substitute" glasses are used. This will be illustrated in the case studies in Chap. 22.

TRANSMITTANCE Most optical glasses transmit light well in the visible and the near IR wavelength spectrum. However, in the near UV, the light is more or less absorbed by most glasses. If an optical system has to transmit UV light, the most commonly used materials are fused silica and fused quartz. Some optical glasses, such as a few SF glasses, have a

reduced transmittance in the deep blue wavelength spectrum, and they have a yellowish appearance. Glass absorption as a function of wavelength is given in glass catalogues for 5- and 25-mm-thick glass plates.

INFLUENCE OF STRESS ON THE REFRACTIVE INDEX Optically isotropic glasses become anisotropic through mechanically and thermally induced stress. This means that the *s* and *p* polarization components of light undergo refraction with different indices of refraction. High index alkali lead silicate glasses (dense flints) display a relatively large absolute refractive index change with a small stress birefringence. On the other hand, borosilicate glasses (crowns) exhibit a small absolute change in refractive index with a relatively large stress birefringence. If the optical system has to transmit polarized light and has to maintain the state of polarization throughout the system or part of it, the choice of materials is very important. For example, when there is a prism in such a system with a relatively large mass, and there is a source of heat in its proximity, there can be a gradient of temperature inside the prism. It will introduce stress birefringence, and the polarization axis will be rotated inside the prism. In this case a better choice for the prism material should be one of the SF glasses rather than crown glasses.

CHEMICAL PROPERTIES Optical glasses acquire their properties through their chemical composition, melting process, and finishing methods. In order to achieve the desired optical properties, optical glasses often exhibit reduced resistance to environmental and chemical influence. There is no single test method sufficient to describe the chemical behavior of all optical glasses. Four characteristics of glass resistance to environmental and chemical influence are given for each type of glass. In the Schott catalogue, glasses are sorted in four groups depending on their climatic resistance, which is the resistance to the influence of water vapor in the air. Water vapor in the air, especially under high relative humidity and high temperature, can cause a change in the glass surface in the form of a cloudy film that generally cannot be wiped off. Glasses are sorted in six groups depending on their stain resistance, which is a resistance to the influence of lightly acidic water without vaporization and possible changes in the glass surface. When the glass is in contact with an acidic aqueous medium, not only can stains appear on the glass surface, but the glass can also be decomposed. Optical glasses are divided into eight groups according to their resistance to acids. The last division of glasses is in four groups according to their resistance to alkalis.

THERMAL PROPERTIES Optical glasses have a positive coefficient of thermal expansion, which means that glasses expand with an increase in temperature. The expansion coefficient, α , lies between $4 \text{ e-}6$ and $16 \text{ e-}6/\text{K}$ for optical glasses. There are a few things that one should consider when designing an optical system to work in the given temperature range:

- Thermal expansion or contraction of glass should not be in conflict with the expansion or contraction of the lens housing.
- The optical system may have to be athermalized, which means that the optical characteristics of the system are unchanged with a change in the lens shape and index of refraction with temperature change.
- Change in temperature can cause temperature gradient in glass, and this can result in temperature-induced stress birefringence.

Most optical design programs have the capability of system optimization simultaneously at several different temperatures. The programs take into account both the expansion of glass elements and changing of their shape, expansion of the housing and the spacers between lenses, as well as the change of index of refraction of the glass materials.

Plastic Optical Materials

In a high-volume production environment, optical components or optical systems require low-cost materials and low-cost fabrication techniques. Plastic optics are used frequently today primarily for this reason. Plastic optical materials also have lighter weight, higher impact resistance, and offer more configuration possibilities than glass materials. Configuration flexibility is one of the greatest advantages of plastic optics. Aspheric lenses and elaborate shapes can be molded, for example, lenses with integral mounting brackets, spacers, and mounting features for easy alignment.

There are some issues, however, that must be considered when using plastic as an optical material. The principal disadvantage of plastic is its relatively low heat tolerance. Plastic melts at a much lower temperature than glass. It is less resistant to surface abrasion and chemicals. Adhesion of coatings on plastic is generally lower than on glass because of the limitation on the temperature at which the coatings are

deposited, due to a low melting temperature of plastic. Further, the durability of coatings on plastic lenses is less robust than on glass. In addition, coatings on plastic often craze over time. The use of ion-assisted deposition of plastic coatings offers harder and more durable coatings on plastic.

The choice of optical plastic materials is very limited, which means that there is not a lot of freedom in the optical design process. A very important limitation is the high thermal coefficient of expansion and a relatively large change in refractive index with temperature. The refractive index of plastic materials *decreases* with temperature (it increases in glasses), and the change is roughly 50 times greater than in glass. The thermal expansion coefficient of plastic is approximately 10 times higher than that of glass. High-quality optical systems can be designed with a combination of glass and plastic lenses. In a combination with glass components in the system, plastic lenses can reduce the price and complexity of the optical system tremendously. When the optical power is mainly distributed over the glass components in the system, with one or two weak-powered plastic aspheric correctors, optical aberrations, especially distortion in wide field-of-view systems, can be very efficiently removed. Weak-powered plastic elements are used to minimize the effect on focus with temperature change.

Plastic optics can be injection molded, compression molded, or fabricated from cast plastic blocks. Fabrication of plastic elements by machining and polishing from cast plastic blocks is economical in the case of large optical elements, where the molding process has severe limitations. Compression molding offers a high degree of accuracy and control of optical parameters. However, injection molding is the most economical process. It offers moderate optical performance, which is acceptable in a lot of applications. Manufacturing of molds is an expensive process, but it pays off in high-volume production. During the system development phase, plastic optics can be very successfully diamond turned for prototyping, since the cost of diamond turning is lower than the cost of the manufacturing of molds. With today's high-quality diamond turning, the scattering effect from the turning grooves is most often under control, and if you have a good vendor, this should not be of concern for visible applications. Sometimes "postpolishing" is required to remove the turning mark residuals.

During the design of systems with plastic elements, the optical designer has to control the shape of the lenses more carefully than for glass elements. The shape (or bending) of the lens should be optimized

for a good flow of the plastic material inside the molds. The thickness of the lens should be quite small, and the parting line, which is the line of contact of two molds, should go through the lens material. It is also important to eliminate inflection points on the lens surfaces in the case of compression molding. This limits the available lens shapes, and requires more parameters to be controlled in the optimization process. Additionally, the lens shape and the refractive index change with temperature have to be monitored, or the system has to be optimized for a given temperature range.

A few of the most commonly used plastic materials are acrylic (poly-methyl methacrylate), polystyrene, polycarbonate, and COC (cycloolefin copolymer):

Acrylic. The most common and important optical plastic material. It has a good clarity and a very good transmission in the visible spectrum, a high Abbe number (55.3), and very good mechanical stability. Acrylic is easy to machine and polish, and it is a good material for injection molding.

Polystyrene. Also a good plastic, cheaper than acrylic, but it has a slightly higher absorption in the deep blue spectrum. Its index of refraction (1.59) is higher than that of acrylic but the Abbe number is lower (30.9). It has a lower resistance to ultraviolet radiation and scratches than acrylic. Acrylic and polystyrene make a viable achromatic pair.

Polycarbonate. More expensive than acrylic, but it has very high impact strength and a very good performance over a broad temperature range. Polycarbonate is often used for plastic eyeglasses. A common form of polycarbonate in eyeglasses is CR39.

COC. A relatively new material in the optics industry, it has many characteristics similar to acrylic. However, its water absorption is much lower and it has a higher heat distortion temperature. COC is also brittle. A new brand name for COC is Zeonex. Comparative properties of optical plastics are shown in Table 6.1.

A Visual Aid to Glass Selection

Figure 6.10 shows what will be referred to as “a visual aid to glass selection.” The glass map is divided into six general regions. The following short summary represents how one selects glasses within these six

TABLE 6.1

Optical and Physical Properties of Optical Plastics

Property	Acrylic	Polystyrene	Polycarbonate	COC
Index @588 nm	1.49	1.59	1.586	1.533
Abbe#	55.3	30.87	29.9	56.2
$d\eta/dT \times 10^{-5}/^{\circ}\text{C}$	-8.5	-12	-10	-9
Linear expansion coefficient/ $^{\circ}\text{C}$	6.5×10^{-5}	6.3×10^{-5}	6.8×10^{-5}	6.5×10^{-5}
Transmission (%)	92	88	90	91
Birefringence	Low	High/low	High/low	Low
Tensile strength (lb/in ²)	10,000	6000	9000	8700
HDT at 264 lb/in ² / $^{\circ}\text{C}$)	92	82	142	120–180
Impact strength (ft-lb/in)	0.3	0.4	>5	0.45
Density (g/cm ³)	1.2	1.05	1.2	1.02
Water absorption (%)	0.3	0.02	0.15	0.01
Advantages	High stiffness, hardness, chemical resistance, and low cost	High index and low cost	Excellent impact resistance and high HDT	High stiffness, high HDT, low water absorption
Disadvantages	Brittle and heat resistant	UV absorption, birefringence, and low-impact strength	High birefringence, low Abbe #, and poor scratch resistance	Brittle

regions. While this is extremely helpful, do note that this does not include glasses with anomalous dispersion characteristics such as the so-called “short flints” of which KZFSN4 is an example (see note below for region 6):

1. Crown (+ elements) low-average dispersion, very common glass
2. Flint (– elements), higher dispersion

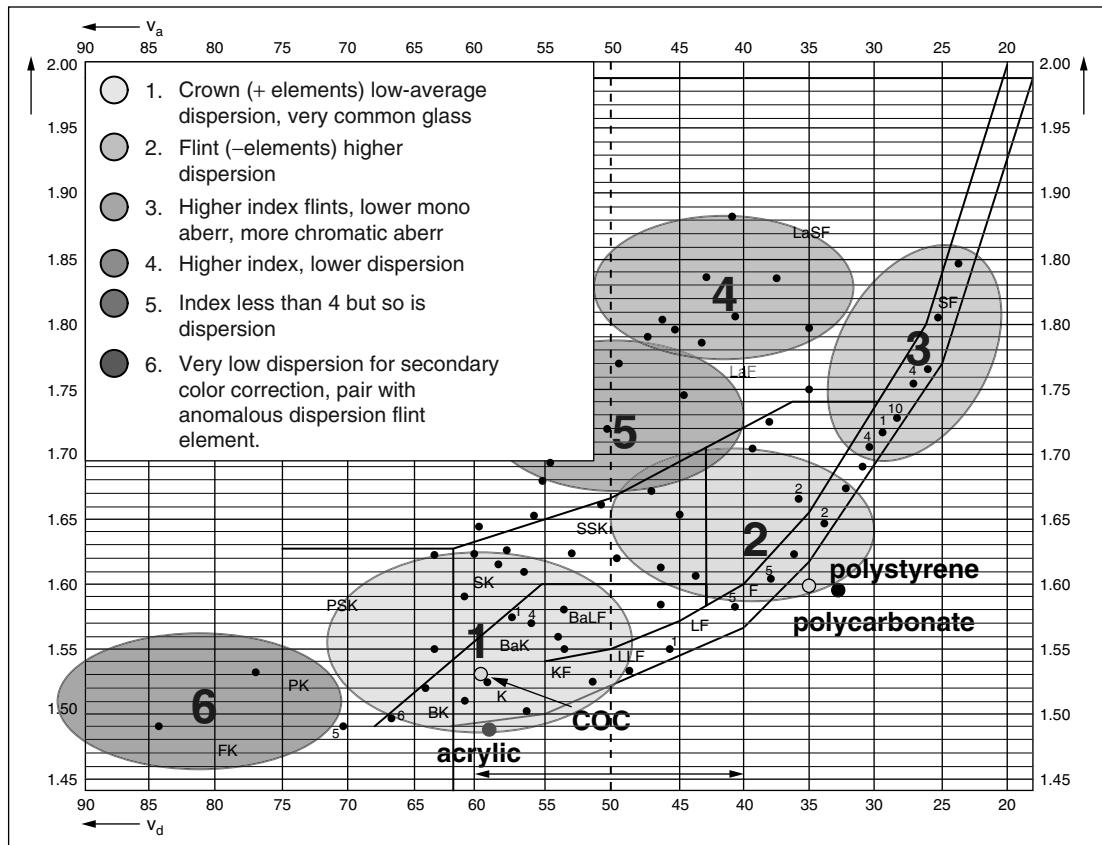


Figure 6.10

A Visual Aid on How to Select Glasses

3. Higher index flints, lower mono aberration, more chromatic aberration
4. Higher index, lower dispersion
5. Index less than 4 but so is dispersion
6. Very low dispersion for secondary color correction, pair with anomalous dispersion flint element

Spherical and Aspheric Surfaces

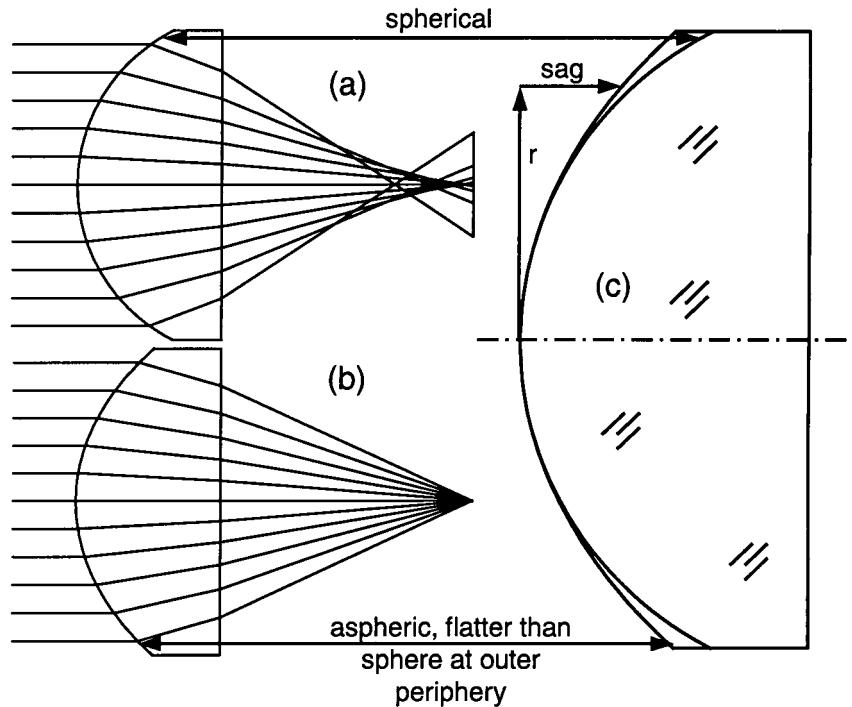
Definition of an Aspheric Surface

A *spherical* surface is defined by only one parameter, the radius or curvature of the surface. If the surface is refractive, with different indices of refraction before and after the surface, then the power of the surface is defined by the surface radius and the indices of refraction of the two media. *Radius* and *curvature* are reciprocal to one another.

Figure 7.1a shows a plano convex lens element with a spherical radius, imaging an axial point from infinity. The spherical aberration is quite evident. The high angle of incidence of the upper limiting ray of approximately 45° to the surface normal causes this ray to refract very strongly and ultimately to cross the axis significantly closer to the lens than rays closer to the optical axis. A spherical surface has the property that the rate of change of the surface slope is exactly the same everywhere on the surface, and thus the aberration is inevitable. Let us consider reducing the slope of the surface toward the outer periphery of the surface in order to flatten the shape in the region surrounding the outer rays. If we make the surface shape gradually flatter as we proceed outward from the optical axis, we can differentially reduce the refracting ray angle so that the net effect is to bring all of the rays to a common focus position, as shown in Fig. 7.1b. Figure 7.1c compares the spherical surface, which is steeper at its edge, with the aspheric surface, which is flatter at its edge. While correction of spherical aberration is not the only application of aspheric surfaces, it is one of the major application areas.

Figure 7.1

Comparison of a Spherical and an Aspheric Lens



Aspheric surfaces cannot be defined with only one curvature over the entire surface because its localized curvature changes across the surface. An aspheric surface is usually defined by an analytical formula, but sometimes it is given in the form of a sag table for coordinate points across the surface. The sag of a surface is shown in Fig. 7.1. The most common form of an aspheric surface is a rotationally symmetric surface with the sag defined as

$$z = \frac{c r^2}{1 + \sqrt{1 - (1+k)c^2 r^2}} + \sum a_i r^{2i}$$

where c is the base curvature at the vertex, k is a conic constant, r is the radial coordinate measured perpendicularly from the optical axis, and $a_i r^{2i}$ are the higher-order aspheric terms.

When an aspheric surface is not rotationally symmetric, it is given either as a *biconic* surface with two basic curvatures and two conic constants in two orthogonal directions or as an *anamorphic asphere*, which has additional higher-order terms in two orthogonal directions.

Another form of aspheric surface is a *toroid* or *toric*. A toroid has, in effect, the shape of a doughnut. If a doughnut were sitting on the table, we all would agree that it had a basic outer diameter. If we now cut the doughnut vertically into two halves and we look at the cut, we see a circle whose diameter is less than the diameter by perhaps a factor of 5 or thereabouts. These two radii define a toroid, the overall outer radius of the doughnut and the smaller cross-sectional radius. If you work with torics in lens design, it is extremely important that you understand fully and completely the definition used by the computer program you are using, so take the time to study the manual in depth in this regard. In addition, if need be, set up a sample to assure that your understanding is correct. While a doughnut is one form of toroid, a football shape is another, and it is imperative to understand which one your equation is representing, especially if it is to be manufactured.

Conic Surfaces

In the case where the higher-order aspheric terms are zero, the aspheric surface takes the form of a rotationally symmetric conic cross section with the sag defined as

$$z = \frac{c r^2}{1 + \sqrt{1 - (1+k)c^2 r^2}}$$

where c is the base curvature at the vertex, k is a conic constant, and r is the radial coordinate of the point on the surface. In Table 7.1, it is shown how a conic surface takes on the following surface types as a function of the conic constant, k , in the sag equation.

TABLE 7.1

Conic Section Types

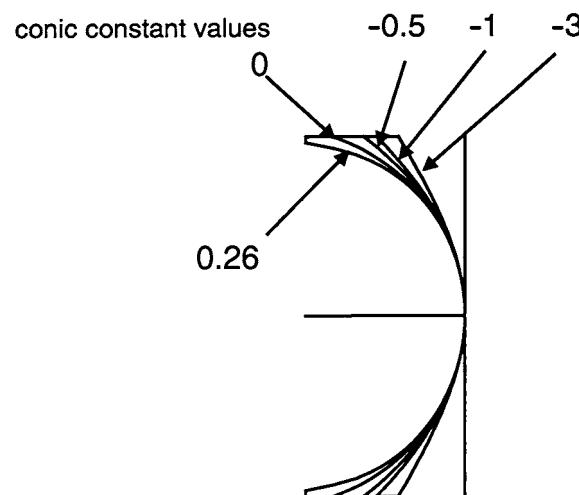
Conic Constant k	Surface Type
0	Sphere
$k < -1$	Hyperboloid
$k = -1$	Paraboloid
$-1 < k < 0$	Ellipsoid
$k > 0$	Oblate ellipsoid

Figure 7.2 shows five surfaces having different conic constants but the same curvature. Most of us are generally familiar with the surface shapes described. One surface we do not come across often is the oblate ellipsoid, sometimes called the oblate spheroid. This can be thought of as the shape of the Earth as it rotates about its axis. Due to centrifugal force, the diameter is greater at the equator than in the polar direction. The oblate ellipsoid has its foci orthogonal to the optical axis.

Conic surfaces, either reflective or refractive, are free of spherical aberration for one particular set of conjugate points. Let us look into a set of different conic surfaces. A spherical surface forms an aberration-free image if the object is at the center of curvature of the surface. An ellipsoid forms an aberration-free image for a pair of real image conjugates on the same side of the surface and a hyperboloid for conjugates on two different sides of the surface. A parabolic mirror forms a perfect image of a point for an axial object at infinity. This is the reason why parabolic mirrors (sometimes combined with hyperbolic mirrors) are widely used in astronomical optics.

When the object point is moved axially from the position of the aberration-free conjugate, a certain amount of spherical aberration is introduced. If the point is moved laterally, other aberrations, such as coma, astigmatism, and field curvature, contribute to image blurring.

Figure 7.2
Conic Surfaces with
the Same Curvature
and Different Conic
Constants



Application of Aspheric Surfaces in Reflective and Refractive Systems

Aspheric surfaces are widely used and often essential in reflective systems due to the small number of surfaces and typically large apertures. While a complex lens may consist of 18 spherical radii in order to minimize the aberrations, a reflective system can only have two surfaces in most cases. A simple spherical reflecting telescope suffers from spherical aberration and coma. A spherical mirror is shown in Fig. 7.3a. A point object at infinity is focused by the spherical mirror at a distance from the mirror equal to the one-half of the mirror radius, and this distance is the focal length. Third-order spherical aberration results, and the wavefront error or OPD is proportional to the $(\text{aperture})^4$, as described in Chap. 5.

A parabolic reflecting telescope is shown in Fig. 7.3b. This is a classic example of how spherical aberration can be corrected. An infinitely distant axial point is imaged to a perfect aberration-free image point. Unfortunately, the image quality degrades quickly when the object is moved off axis. Coma is the aberration that restricts the field of view of the parabolic telescope to a very small field.

A very common form of reflecting telescope is the two-mirror Cassegrain telescope, with both mirrors being conic surfaces. The *classical Cassegrain telescope* has a paraboloidal primary mirror and a hyperboloidal secondary mirror, as shown in Fig. 7.4. f_1 is the location of the

Figure 7.3
Conic Surfaces for
Reflecting Systems

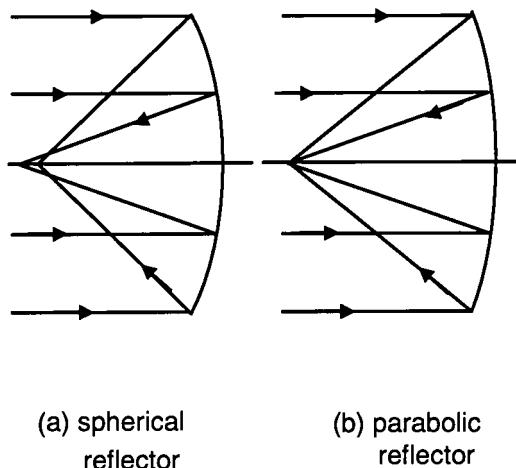


Figure 7.4
Cassegrain Telescope

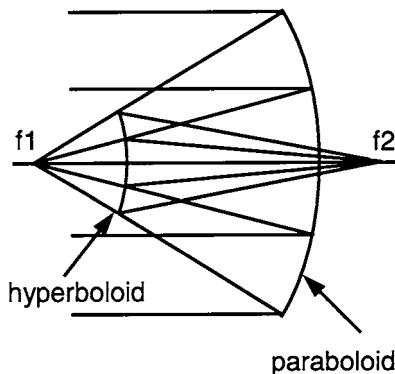
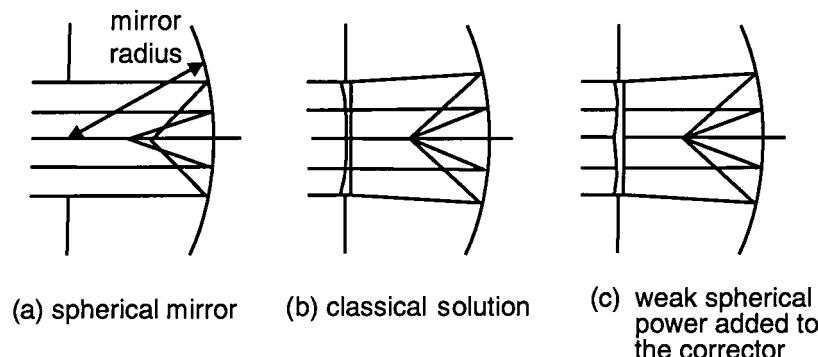


image which would be formed by the large primary mirror, and f_2 is the location of the image of the entire system. Note that the secondary mirror reimages f_1 to f_2 . A similar configuration, called a *Ritchey-Chrétien* Cassegrain telescope, has both primary and the secondary hyperbolic mirrors. The classical Cassegrain performance is limited by off-axis coma while the *Ritchey-Chrétien* Cassegrain is, in effect, a coma-free Cassegrain. Its limiting aberration is astigmatism.

Another well-known type of telescope is a Schmidt telescope, which is shown in Fig. 7.5. It consists of a spherical mirror with an aspheric corrector plate located at the center of curvature of the mirror. Third-order spherical aberration results in a wavefront aberration function proportional to the (aperture)⁴. If the aperture stop is located at the center of curvature of the mirror, there is symmetry for all field positions. Apart from the aperture stop being obliquely viewed by the

Figure 7.5
Schmidt Telescope



oblique bundle of off-axis rays, the oblique rays are focused in the same manner as the axial bundle. The chief ray is normally incident onto the mirror everywhere in the field of view. The image is formed on a spherical surface, with the image radius equal to one-half the mirror radius.

Without any aberration correction, the rays that are closer to the aperture edge are focused closer to the mirror than the paraxial rays. The wavefront distortion, which is proportional to the fourth power of the aperture radius, can be corrected with a wavefront distortion of the opposite sign introduced by the aspheric corrector plate placed in the aperture stop as in Fig. 7.5b to provide effective “parabolization” of the spherical mirror. A fourth-order aspheric deviation from the flat base surface of the glass corrector introduces a negative fourth-order wavefront distortion. The aspheric refractive corrector reduces the spherical aberration. However, some chromatic aberration is introduced by the wedged shape at the outer periphery of the corrector, close to the aperture edge. In order to minimize this chromatic aberration, a very weak positive power is added to the corrector, such that the corrector has zero power at 0.7 of its aperture, as shown in Fig. 7.5c. This shape of the corrector is not only the optimum shape for the correction of spherico-chromatism (the variation of spherical aberration with wavelength), it is also best suitable for manufacturing.

The majority of optical systems are based on the use of spherical components because they are easier to manufacture. However, there are cases where aspheric optical components have a significant advantage over spherical ones. In astronomical optics, reflective aspheric components are widely used. Today, with the significant development of new plastic materials, low-cost molded aspheric refractive optics are finding their place in the large consumer market. Precision diamond grinding and compression molding of glass aspheric lenses is also becoming more common.

Refractive aspheric lenses are widely used in many kinds of illumination systems, from the condensers in projection systems and microscopes, to street lamps and searchlights. Since in many cases these are not imaging systems, the manufacturing tolerances on these components are somewhat forgiving. In the case where the optical components are not exposed to the heat from the light source, aspheric optical elements can be injection molded. In projection systems, aspheric condensers are often molded glass lenses, mostly from B270 glass. B270 is a very common low-cost glass similar to BK7, which is used extensively in “float glass” for low-cost windows and mirrors. Glasses that are moldable have a

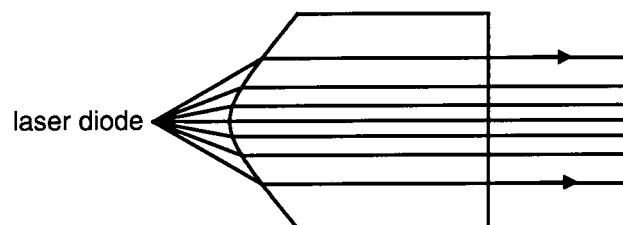
lower temperature at which they become soft than the standard optical glasses. Heat generated by the light source, or by the absorption of light by the optical components themselves, is often a severe problem in optical systems and requires the use of heat-resistant materials such as fused silica or fused quartz.

Another field where aspherics find their place is in systems for focusing of laser beams, for example, in CD players or data storage systems, when coupling light from a laser diode into a fiber, or when collimating light from laser diode arrays. These applications require high-precision optical components, as well as optically stable components in a given temperature range. The small size of these components makes them easier to mold. Optical plastics are used wherever they are acceptable, because of the much lower manufacturing costs. However, accurate lens shape and the very good temperature stability of glass aspheric lenses make them a better solution for applications where high precision is required. Glass aspheric lenses are manufacturable in diameters smaller than 25 mm. At diameters greater than 25 mm, aspheric glass lenses become too expensive for high-volume manufacturing. Single-glass bi-aspheric lenses are used for focusing or collimation of $NA = 0.5$ laser beams with diffraction-limited performance.

An infinitely distant object imaged through a planohyperbolic lens is focused to an aberration-free spot. This feature is used in the case of a planohyperbolic fiber lens (the cross section of the fiber has a planohyperbolic shape) to collimate the fast axis of the laser diode arrays, as shown in Fig. 7.6. The distance of the laser diode from the lens is determined by the index of refraction of the lens.

Improvements in the quality of injection-molded plastic optics make the use of them possible in camera lenses and projection lenses. The optical design of these kinds of lenses is difficult because many parameters have to be considered such as

Figure 7.6
A Planohyperbolic Collimator



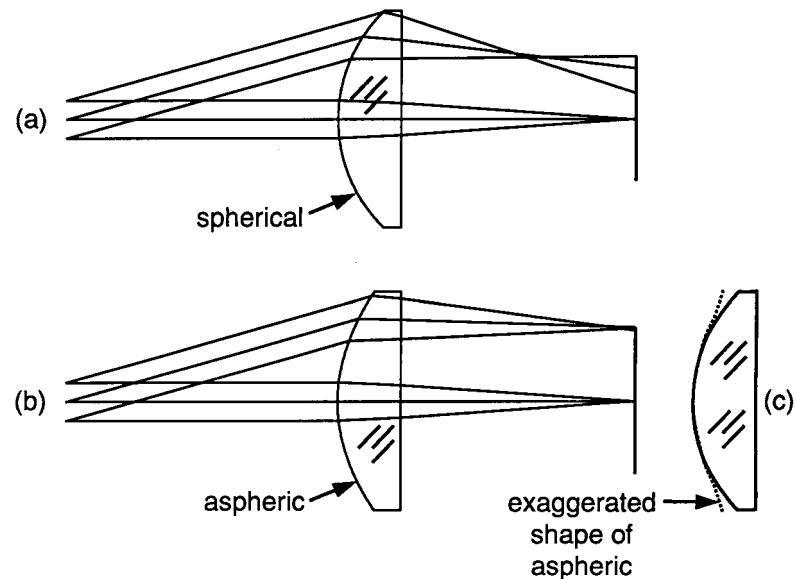
- The location of the component in the lens to minimize the beam size over the plastic component.
- The shape of the lens to keep the molding parting line inside the component.
- Athermalization of the lens.

In high-performance lenses, plastic components should be away from the aperture stop because it is very difficult to achieve diffraction-limited performance with injection-molded plastic components. However, they can be extremely useful in the correction of field aberrations such as astigmatism, field curvature, and also distortion.

Much of the discussion thus far with respect to aspheric surfaces and their benefits has related to the correction of spherical aberration. If an aspheric surface is located at or near the aperture stop of a system, it will primarily affect or benefit spherical aberration, which is an axial aberration which, for the most part, carries across the field of view. As aspheric surfaces are located further from the stop they can help to minimize some or all of the off-axis aberrations such as coma and astigmatism.

A good example of the application of an aspheric surface used for astigmatism correction is shown in Fig. 7.7. In Fig. 7.7a, we see a single-element lens with its aperture stop located far to the left of the lens. If

Figure 7.7
Correction of Astigmatism with an Aspheric Surface



the curved lens surface is spherical, the oblique rays create a footprint on the surface, which is larger in the plane of the figure than the orthogonal plane in/out of the figure. As we discussed in Chap. 5, this tends to refract and pull the rays in the plane of the figure inward from where they would otherwise focus. We now need to ask ourselves what would it take to push the focus position outward and compensate for this inward focus shift. The answer is to create a more negatively powered surface in the plane of the figure at the outer periphery of the lens. This is shown in proper scale in Fig. 7.7b and in an exaggerated form in Fig. 7.7c. This more negatively powered surface shape in the plane of the figure has virtually no effect in the orthogonal plane, hence the highly efficient correction of astigmatism by the aspheric surface.

Another common use of aspherics is in the thermal infrared where the cost of materials is extremely high. With the use of aspherics, the number of elements can be reduced to a minimum.

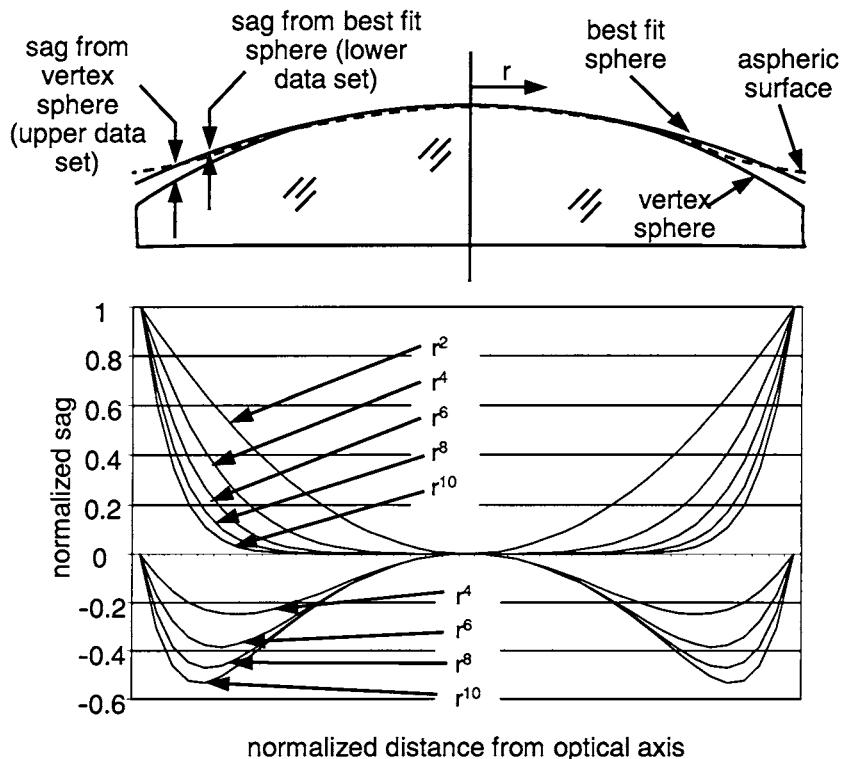
Guidelines in the Use of Aspheric Surfaces

The proper usage of aspheric surfaces is extremely important. This includes which surfaces to make aspheric and whether to use a conic section or, alternatively, a higher-order aspheric. The conic sections include paraboloids, hyperboloids, and ellipsoids, as discussed earlier in this chapter. The higher-order terms are surface departures from conic, which are proportional to r^4 , r^6 , r^8 , r^{10} , and so on, where r is the radial distance from the optical axis. The simpler forms of reflective systems, such as the classical Cassegrain (paraboloidal primary, hyperboloidal secondary) and the Ritchey-Chrétien Cassegrain (two hyperboloids), were discussed earlier. The classical Cassegrain is limited by coma and field curvature, and the Ritchey-Chrétien is, in effect, a coma-free Cassegrain which is limited by astigmatism and field curvature. Once the basic system is set up on your computer, varying the appropriate conic constants is all it takes to reach a viable solution.

But how do we decide which surface, or surfaces, in a lens system should be made aspheric, and how do we decide what form of aspheric to use? To answer this question, first consider Fig. 7.8, where we show: (1) the aspheric surface departures from the base spherical surface and (2) the departures from what we call the “nearest sphere” or “best-fit

Figure 7.8

Aspheric Sags from Vertex Sphere (Top) and Best-Fit Sphere (Bottom)



sphere" to the aspheric surface. The nomenclature is shown in the top of Fig. 7.8 with accurate data below.

If we now compute for our baseline spherical optical system a plot of the optical path difference, we should look for a form matching the basic profile or character of these data. For example, if the axial OPD plot resembles the form of the sag from the nearest sphere for the r^6 case, then varying the r^6 coefficient on a surface near the aperture stop will likely be beneficial. If we find a sharp increase or decrease in the OPD off axis at the edge of the pupil, then varying a higher-order term or two on a surface away from the stop will likely be beneficial. There are some basic guidelines, and these are listed here:

1. Conic surfaces can be used for correcting third-order spherical aberration and other low-order aberrations.
2. If you have a nearly flat surface, then use an r^4 and higher-order terms rather than a conic.

3. If you have at least a somewhat curved surface, then you can use the conic along with higher-order terms if required.
4. It is generally best not to use both a conic and an r^4 surface, as they are mathematically quite similar. This is because the first term of the expansion of a conic is r^4 . While they can both literally be used, the optimization process often tends to beat one against the other, yielding artificially large coefficients, and this may have an effect on the convergence of the optimization.
5. Use aspherics beginning with the lower-order terms and working upward as required. If you can stay with conics, this may make testing more manageable. You should be able to assess the need for adding terms based on the character of the OPD plot.
6. It is very dangerous to use a large number of aspheric surfaces, especially with higher-order terms. This is because they will beat against each other. This means that as one surface adopts a certain aspheric profile or contour, it may increase in its asphericity, with its effect cancelled by adjacent surfaces. For example, if the first of two closely located aspheric surfaces has significant surface departure from sphericity, the neighboring aspheric surface could very likely cancel this effect. While the lens may perform well on paper, we now need to manufacture two highly aspheric surfaces, a difficult and expensive task which may not be necessary.
7. If possible, optimize your design first using spherical surfaces, and then use the conic and/or aspheric coefficients in the final stages of optimization. This may help in keeping the asphericities to a more manageable level.

Specification of Aspheric Surfaces

It is important to specify an aspheric surface sufficiently enough to convey to the shop both what you want and what you need. The following items are most often included in specifying aspheric surfaces:

1. The surface to be aspheric is labeled aspheric on the component drawing.
2. You should include an equation of the surface shape along with the aspheric coefficients. A small sketch indicating the nomenclature and sign convention is recommended.

3. A table listing the sag as a function of the radial distance from the surface vertex normal to the optical axis, r , is imperative. You should list a sufficient number of data points to adequately sample the surface profile.
4. You should list how close the actual surface must come to the ideal design prescription. The form of this can be “surface to match nominal surface to within four visible fringes (or 0.001 mm) over clear aperture.”
5. You may need to call out higher-frequency surface irregularities and/or surface finish. The higher-frequency irregularities can be called out by indicating the maximum slope departure from nominal over the surface. Surface finish is normally called out by indicating the rms surface finish, in nanometers. This latter callout is generally used for diamond-turned surfaces, where surface roughness is sometimes a problem or where scattering and off-axis rejection is of major concern such as in space telescopes.
6. You should, if possible, indicate the form of testing to be used.

Do keep in mind that the more callouts you list and the more extensive the testing, the more costly your optics will be, and they will likely take longer to manufacture. Your callouts should indicate only what you need functionally for your system to work properly.

Design Forms

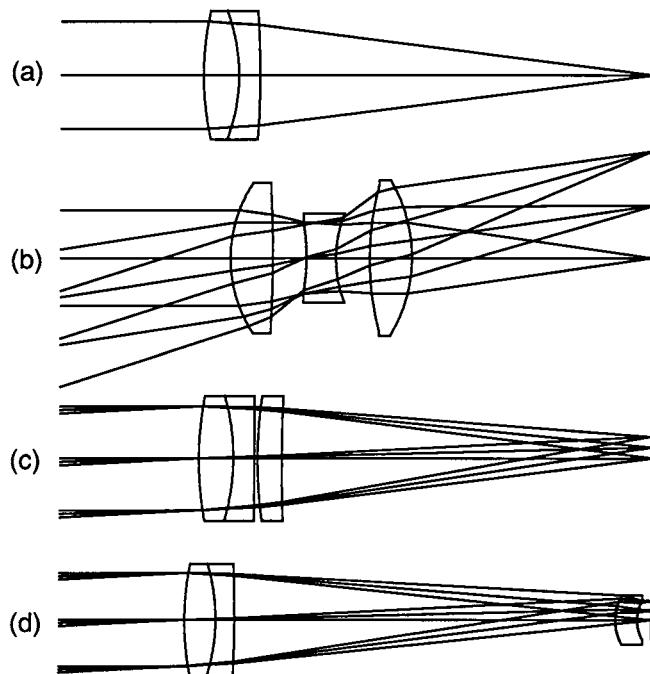
Introduction

In this chapter, we will discuss how we select the proper design form or configuration for both refractive and reflective image-forming systems. We will also consider fold mirrors and prisms since they have a significant influence on the system design configuration.

The proper system design form or “configuration” of an optical system is generally the key to a successful design effort. The term “configuration” here means the basic form of the system which includes not only the number of elements, but also the relative optical power and distribution of the elements within the lens system. For example, an achromatic doublet of two cemented elements, as shown in Fig. 8.1a is clearly different in form from a Cooke triplet, which consists of three separated elements, as in Fig. 8.1b, with two outer positive crown elements and a negative flint element at the center. The Cooke triplet can be used over wider fields of view than a doublet due largely to a reasonable degree of symmetry fore and aft of the central element, which is at or near the aperture stop. The doublet and triplet are very different configurations.

What if we were to add a single positively powered element immediately following a cemented doublet, as in Fig. 8.1c? Would the lens configuration be called a triplet? It certainly would not be a Cooke triplet as the symmetry is not present. This is a very different configuration or design form from a Cooke triplet. However, it is, of course, a three-element lens. The same is true for Fig. 8.1d, where we again have three elements, only here the third element is very near to the image and is serving to both flatten the field as well as correct astigmatism. All three of the

Figure 8.1
Doublets and Triplets



three-element configurations are quite different in configuration or form, each with their relative advantages.

The selection of an optimum configuration prior to initiating a design effort provides the starting point from which the design optimization proceeds. While lens design software has improved significantly over the years, the programs are rarely capable of changing configurations, and never add or delete elements. Most of the time the program will reach an optimum or local minimum in the error function for the input configuration. For more information on the optimization process, see Chap. 9. The configuration selection is driven by many factors. Nearly every system specification can have an influence on the configuration. The major factors influencing the configuration selection are:

- Field of view
- Performance requirement
- $f/ \#$
- Packaging requirements
- Spectral range

System Configurations for Refractive Systems

We will review a progression of configurations for lenses in order to illustrate just what differentiates one from another. The following configuration forms are shown in Fig. 8.2.

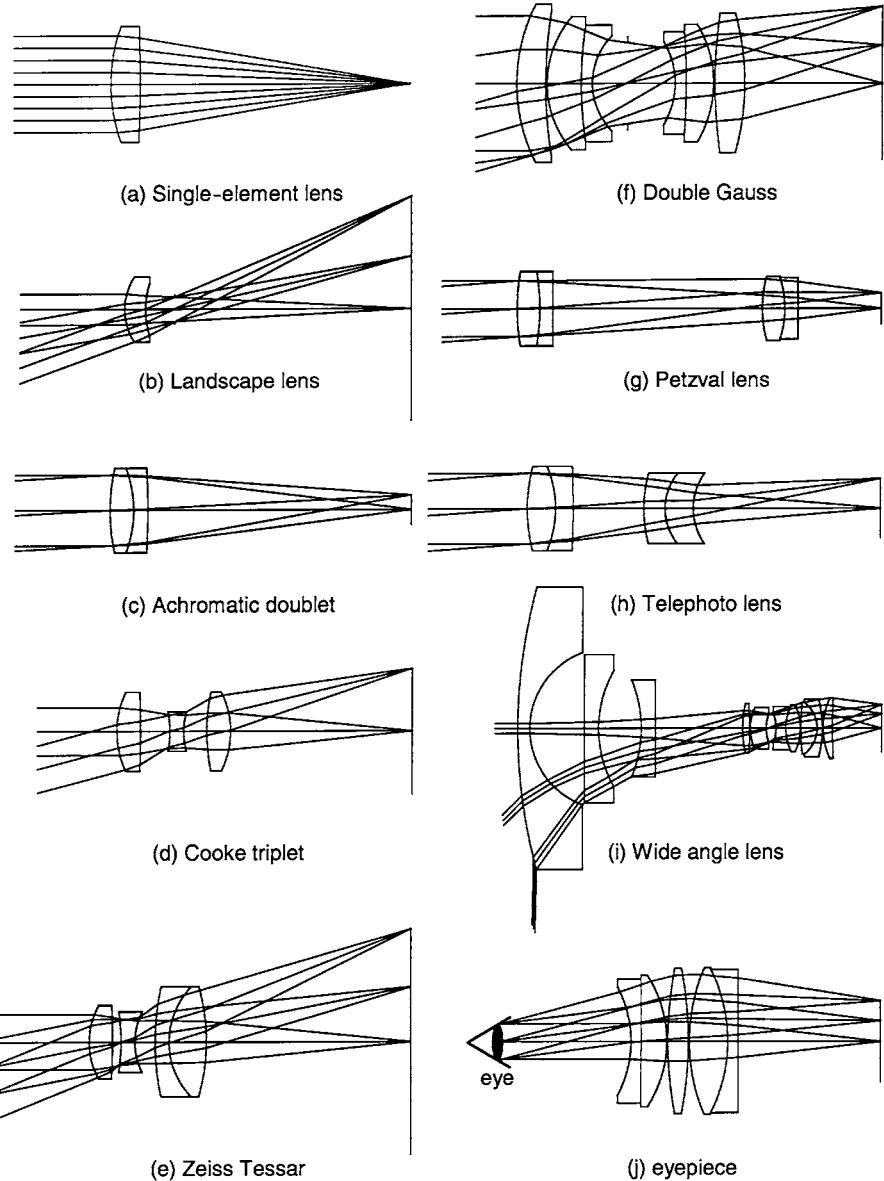
SINGLE-ELEMENT LENS (FIG. 8.2a) A single-element lens has generally poor image quality and a very small field of view. Further, it suffers from chromatic aberrations and it can only be used at a high $f/ \#$. We often think of a single element as of “magnifying glass quality.”

It is important to note that the performance, which may be poor for one application, may be just fine for another. For example, we have stated that a single element generally has poor image quality. This is, for the most part, true for most critical imaging applications such as camera lenses, machine vision optics, and other similar applications. However, if you are looking for a photon collector with little or no image quality requirements, then a single element may be quite adequate for the task. Another good example is the optics used for optical data storage and other microoptics applications. For data storage applications, a laser diode is imaged to a micron or submicron spot diameter. Since the laser is nearly monochromatic (there may be thermally induced shifts in wavelength), the field of view is nearly zero and the scale or size of the system is extremely small, we often find that a single aspheric element is sufficient for the task.

LANDSCAPE LENS (FIG. 8.2b) While a landscape lens is also a single element, it has an aperture stop which is remote or separated from the lens itself. Further, the lens is bent or “curled” around the stop for symmetry reasons. This reduces the angles of incidence on the surfaces and thereby reduces off-axis aberrations. Earlier in Chaps. 3 and 5 we discussed in greater depth how minimizing angles of incidence within a lens system reduces aberrations. A landscape lens can have its aperture stop either aft of the element as shown or in front of the element. It can be shown that the aberrations are somewhat reduced if the stop is in front of the lens, and many early box cameras were constructed this way. The landscape lens has chromatic aberration as well as residuals of many of the other third-order aberrations.

Figure 8.2

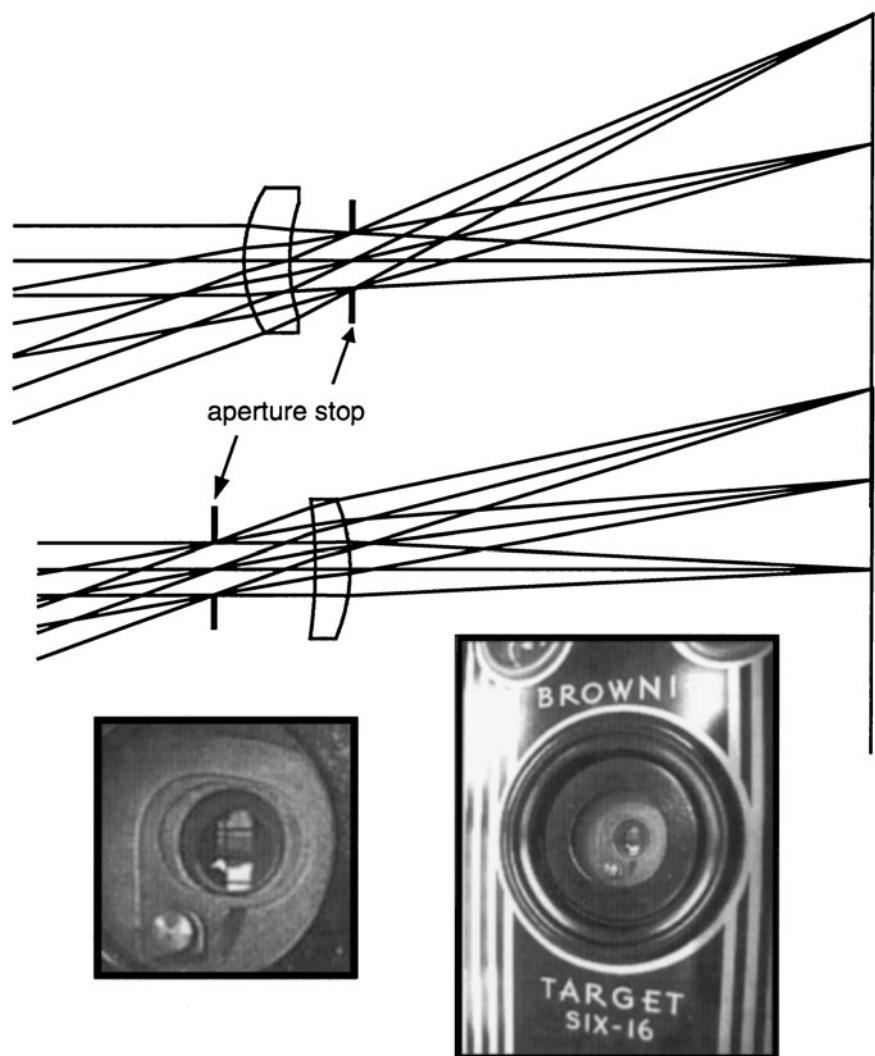
Progression of Configurations



We show in Fig. 8.3 two forms of landscape lenses, one with the stop aft of the lens and the other with the stop forward of the lens. It can be shown that the performance is slightly improved with the stop in front. Also shown are two photos of the front of an early Kodak Brownie camera with a flat window followed by its aperture stop and finally the lens, which is aft of the stop.

Figure 8.3

Landscape Lens with Stop Aft and Forward of Lens



ACHROMATIC DOUBLET (FIG. 8.2c) The achromatic doublet is capable of bringing the red and blue wavelengths to a common focus, with the central green or yellow wavelength defocused slightly toward the lens. A typical achromatic doublet has a blur diameter approximately 25 times smaller than an equivalent single-element lens (based on an f/5 lens in the visible spectral band).

A cemented doublet performs well only over a small field of view, and it cannot be used at low $f/\#$ s due to higher orders of spherical aberration. In order to balance the inherent third-order spherical aberration of the cemented doublet, one can introduce a small airspace between the elements. This airspace will permit the balancing of fifth-order spherical aberration with the inherent third-order aberration for an improved overall level of performance. Further improvement can be realized by adding an additional element near the image, which can be used as a field flattener, and often it is possible to bend this element to balance and eliminate some or most of the astigmatism.

COOKE TRIPLET (FIG. 8.2d) This three-element lens form takes advantage of symmetry in order to minimize the angles of incidence of the rays as they proceed through the lens over the field of view, and hence it is capable of an acceptable level of performance for many applications. The Cooke triplet was first designed in England by H. D. Taylor at the "Cooke and Sons" optical company. The Cooke triplet is the first configuration we have presented in this review that allows for the optimization and balancing of the seven primary, or third-order, aberrations, as well as the control of focal length. There are eight "useful" variables in a Cooke triplet, the six radii and the two airspaces. Element center thicknesses are generally not of significant use in aberration control and, for the most part, element thicknesses that yield reasonable manufacturing ease are best. We can thus control or optimize the following:

1. Spherical aberration
2. Coma
3. Astigmatism
4. Axial color
5. Lateral color
6. Distortion
7. Field curvature
8. Control of the focal length

It is important to note that just because we have the same number of useful variables as the number of primary aberrations (along with the focal length), this does *not* at all mean that the aberrations can be brought to zero or even close to zero. What it means is that for the $f/\#$ and focal length is selected, the aberrations can be reasonably well balanced against one another, especially the third-order aberrations. Thus, for example, an $f/6$ Cooke triplet covering a 10° full field of view over the visible spectrum will likely be capable of a reasonable level of performance. However, an $f/1.4$ Cooke triplet covering a 30° full field of view will probably provide fair to poor performance, at best. The low $f/\#$ will lead to significant spherical aberration residuals, and the wide field of view will lead to coma, astigmatism, and other off-axis aberrations.

ZEISS TESSAR (FIG. 8.2e) The Zeiss Tessar is derived from, and is an improvement upon, the Cooke triplet. Paul Rudolph of Zeiss Jena replaced the original single rear lens in the Cooke triplet with a doublet lens, resulting in a better lens performance, with higher resolution, excellent contrast, and very low levels of distortion.

There is a rule of thumb in lens design which says “clip it in the bud.” What is meant here is that the best place to correct or eliminate aberrations is as close to where they are being introduced as possible. In the case of a Cooke triplet we see that the first positive lens element takes collimated light from infinity and bends or converges it into the second negative lens element. The second element takes the slightly converging light and diverges it into the third and positive element. And finally, the third positive element takes the slightly diverging light and bends the rays so as to create the required $f/\#$ of the lens. The amount of ray bending or redirection is greatest for the third element than for the first or second elements. Thus, the aberrations introduced by the third element will be greatest as well. This makes the third element an excellent candidate to convert to a doublet. As one of the more difficult aberrations to correct in a Cooke triplet is axial color (change in focal length with wavelength), making the third element a doublet (as in the Tessar) allows for a superior level of correction of this as well as other residual aberrations. Tessar designs can be effectively used at $f/\#$ s down to $f/4.5$ or somewhat lower, depending on the relative mix and level of lens requirements. Further, as will become apparent, the Cooke triplet forms the basis of many more complex and high-performance configurations.

DOUBLE GAUSS (FIG. 8.2f) The double Gauss lens is yet a further extension toward improved performance at lower $f/\#$ s and wider fields of view. If we summarize what we have learned thus far, we can see more clearly the evolution of the double Gauss lens. The methodologies learned thus far include *splitting optical power* to minimize aberrations, using *negatively powered elements with smaller beam diameters* for field curvature correction as in the Cooke triplet, and using *symmetry fore and aft of the aperture stop* also for symmetry reasons, thereby minimizing the angles of incidence on the lens surfaces. As we will learn later, symmetry also allows for cancellation of several off-axis aberrations. The double Gauss lens uses at least two negatively powered elements near the stop and two or more positive elements on the outsides. Furthermore, there is a reasonably high level of symmetry surrounding the aperture stop.

The double Gauss lens is capable of good performance down to about $f/1.4$ and even lower. Indeed, there have been several $f/1.0$ double Gauss lenses in the 35-mm camera marketplace, including the famous *Nocitlux* designed by Walter Mandler of Ernst Leitz Canada. As with any lens, there are compromises, and at $f/1.0$ the lens is hardly diffraction limited; however, from an overall performance standpoint and light-gathering capability, the lens is a top performer. We should note that for the more demanding levels of performance (such as the Nocitlux), higher refractive index materials are often used as they can significantly reduce aberrations, as well as anomalous dispersion glasses for superior chromatic aberration correction. Glass selection is discussed in Chap. 6.

PETZVAL LENS (FIG. 8.2g) The Petzval lens represents a very different design philosophy. This lens is intended for smaller fields of view and only moderate $f/\#$ s such as $f/3.5$ or slower. The design philosophy here is to use two separated doublets with the power task shared between the two. This yields lower secondary chromatic aberrations than a single doublet of the same net $f/\#$. This form of design is used for high-performance small field-of-view lenses as one might encounter in aerial reconnaissance, for example. As noted, the Petzval lens is not well suited for wide-angle applications, as there is little opportunity for symmetry as with the Cooke triplet or the double Gauss.

TELEPHOTO LENS (FIG. 8.2h) The telephoto lens is a positive group of elements, followed by and separated from a negative group of

elements. As we showed earlier in Fig. 8.1, this form of lens has a focal length longer than the physical length of the lens, hence the name “telephoto.” The ratio of the physical length to the focal length is called the *telephoto ratio*. In the case of a fast lens (low *f*/number), with a low telephoto ratio below 0.6, for example, the lens configuration becomes quite complex. In the limit, if the light exits the second group collimated, we have a beam contractor or, in effect, a Galilean telescope, and the focal length is infinite. It is important to note that both the positive and the negative groups generally need to be separately achromatized in order to produce a complete lens with sufficiently low chromatic aberration.

It is interesting to think about the use of the words “telephoto lens.” To a photographer a telephoto lens is generally a lens whose focal length is longer than a standard lens for a specific film format. For example, the standard focal length for a 35-mm camera lens is in the order of 50 to 55 mm. A lens with a 100- to 135-mm focal length or longer is generally considered to be a telephoto lens. This is because objects appear closer due to the smaller field of view covered by the longer focal length lens. An achromatic doublet with a focal length of 135 mm might, to a photographer, also be a telephoto lens because it brings the object closer. However, to a lens designer a telephoto must be of a form or configuration as shown using a positively powered front group and a negatively powered rear group so that the focal length is longer than the physical length of the lens.

WIDE-ANGLE LENS (FIG. 8.2*i*) A lens covering a substantially wider field of view than a normal lens (in photography, for example) is called a *wide-angle lens*. Thus with a standard focal length of 50 to 55 mm in 35-mm photography, a wide-angle lens is generally considered to be 35 mm or less. In order to cover the wider fields of view, we often use a strong negatively powered front element or group of elements to bend the rays outward to cover the wider field angles. In order to still image from infinity, the light in the space between the main body of the lens and the front negatively powered group needs to be converging toward the object as seen in Fig. 8.2*i*.

We show the wide-angle lens as having a negatively powered three-element front group with a multielement configuration for the prime lens group. What is happening here is that the prime lens group, itself, is covering a smaller field of view, with the field angle increase happening only at the negatively powered front elements.

EYEPIECE (FIG. 8.2j) An eyepiece is used to visually view and magnify the image from a microscope objective or a telescope objective, or alternatively to view a display such as in a head-mounted display system. The eyepiece is a very different configuration of lens system in that the aperture stop is not only quite remote from the main part of the lens, but it is in reality the pupil of the eye. Eyepieces are normally designed by tracing rays from the eye, which is the aperture stop, to the image plane, as shown in Fig. 8.2j.

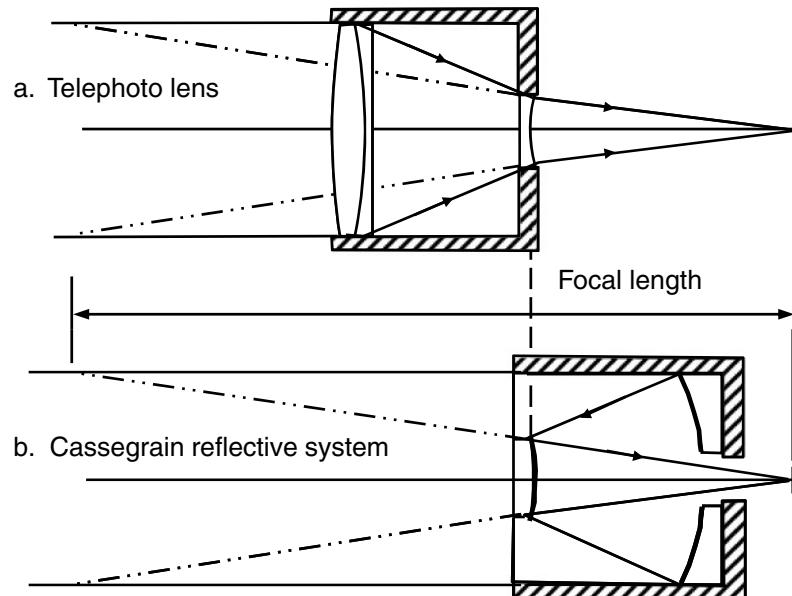
As can be seen in Fig. 8.2j the rays at the extreme field of view are primarily using the outer periphery of the lens elements. Unfortunately, this often results in significant amounts of astigmatism, lateral color, coma, and distortion. These field aberrations can be quite significant and difficult if not impossible to correct for in extremely wide field-of-view eyepieces with 60° field of view or more. The careful use of higher-index glasses combined with one or more aspheric surfaces (if possible) can help to mitigate the problems to some extent. Eyepieces represent a very different configuration form than other lenses discussed thus far.

System Configurations for Reflective Systems

As with lenses, reflective or mirror systems can be of many varied configurations. There are, however, some inherent and very basic differences between refractive or lens systems and reflective or mirror systems. Lens systems are most often straight through and use the full clear aperture of the entrance pupil, as shown in the simple telephoto lens example in Fig. 8.4a. Mirrors have the fundamental challenge that they get in each other's way as shown in the two-mirror Cassegrain system in Fig. 8.4b. The Cassegrain is the reflective analogy of the telephoto lens described earlier in this chapter. Note that the large objective lens shown as an achromatic doublet is the optical analogy of the large concave mirror generally called the *primary mirror*. Further, the smaller lens in the refracting telephoto lens is the analogy of the small convex mirror in reflective systems, and it is generally called the *secondary mirror*. Note that both lens and mirror systems are in the form of the telephoto lens and the focal lengths are in fact equal in both systems.



Figure 8.4
Telephoto and
Cassegrain
Configurations



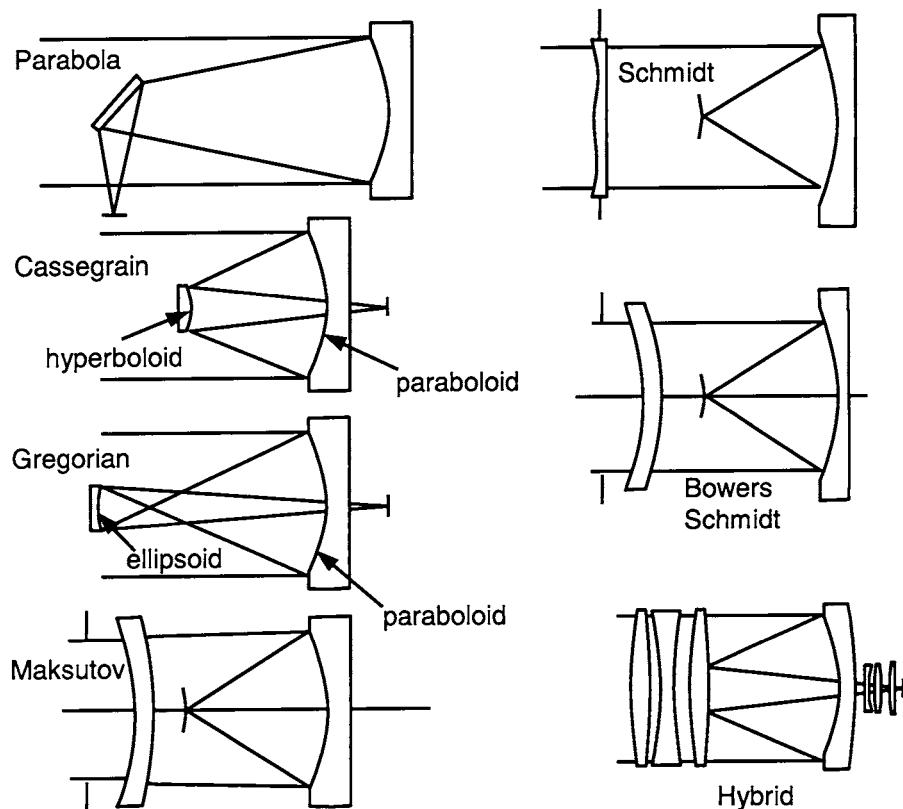
The fundamental issue of the mirrors getting in each other's way in reflective systems leads to many significant differences between reflective and refractive systems.

We will review a progression of configurations for reflective systems in order to illustrate just what differentiates one from another. The following configuration forms are shown in Fig. 8.5.

PARABOLOID A single parabolic mirror has zero spherical aberration on axis for an infinitely distant object. It is, however, limited by coma off axis. As shown, the image-forming light is often folded out to the side via a tilted flat-fold mirror sometimes called a diagonal mirror. In this case it is called a Newtonian telescope.

CASSEGRAIN The Cassegrain form of reflective system is perhaps the most common. In its "classical" design form, a parabolic primary mirror and a hyperbolic secondary are used. With this prescription, coma is the limiting aberration and is the same as a single parabolic mirror of the same f /number. An improved level of performance is achieved by allowing the primary mirror to be hyperbolic along with a hyperbolic secondary. This solution is called the Ritchey-Chrétien form

Figure 8.5
Reflective
Configurations



of Cassegrain. It is, in effect, a coma-free Cassegrain and is limited only by astigmatism and field curvature. The layout itself is virtually identical for the classical Cassegrain and the Ritchey-Chrétien forms. In most configurations, the Cassegrain has an inward or concave curving image field due to the Petzval contribution of the convex secondary mirror predominating over the concave primary mirror.

GREGORIAN A concave parabolic primary mirror with a concave elliptical secondary mirror is the Gregorian form of telescope. In effect, the elliptical secondary mirror reimages the image formed by the primary mirror to its final position aft of the primary as shown in Fig. 8.5. The Gregorian is not as common as it was some years ago, and the reason for this is that astronomers years ago did not believe that convex aspheric mirrors (as required by the Cassegrain) could be effectively

fabricated and tested. With the advent of interferometry and other advances in optical metrology, testing convex aspheric surfaces became more viable and the more compact Cassegrain is now more widely used.

MAKSUTOV The Maksutov uses a spherical primary mirror with a spherical weakly powered meniscus glass corrector plate to balance the spherical aberration of the mirror. This system has been popular with amateur astronomers for many years. A variation known as the *Maksutov-Cassegrain* locates the corrector closer to the primary, and an aluminized spot at the center of the convex surface acts like a Cassegrain secondary mirror, hence the name Maksutov-Cassegrain. This is the design form used in the well-known Questar telescope for many years.

SCHMIDT The Schmidt system uses a thin aspheric corrector plate located at the center of curvature of a spherical primary mirror to effectively correct all orders of spherical aberration. The aperture stop is at the corrector plate, and is located at the center of curvature of the primary mirror. Due to this geometry, the chief rays at all field angles will be incident onto the primary perpendicular to its surface. To third order, the spherical aberration will be eliminated at all field angles. The image is formed on a spherical image whose radius is equal to the focal length of the primary mirror. Further, the Schmidt system works quite well at low *f*/numbers, even as low as *f*/1 or less.

The fact that most aberrations are zero is due to what is known as the "Schmidt principal," which has as its basis the aperture stop being located at the center of curvature of a spherical mirror. It can be shown that in addition to zero spherical aberration, there is no third-order coma, astigmatism, or distortion. The major residual aberration of tangential oblique spherical aberration is due to the ray obliquity on the corrector plate as well as the foreshortening of the entrance pupil at off-axis field angles. The Schmidt principal is an exceptionally powerful technique, which has been successfully applied in the design of many well-corrected optical systems. It is especially useful in wide-angle applications. A variation of the Schmidt, invented by Baker, uses three separated aspheric Schmidt plates for minimization of the off-axis residual aberrations.

There is a variation of the Schmidt system known as a *shortened Schmidt Cassegrain*. In this configuration, we move the corrector plate closer to the primary mirror to a location where the ray bundle diameter is similar to what it would be in a Cassegrain system. Then a convex mirror is mounted on the interior of the corrector plate and the system

geometry now resembles a Cassegrain. The fact that the corrector plate and the aperture stop are no longer at the center of curvature of the spherical primary mirror are cause for coma and other off-axis aberrations; however, the overall performance is reasonably good for small fields of view. This form of system is extremely popular in contemporary amateur telescopes due to its robust performance combined with aggressive packaging.

BOWERS SCHMIDT This configuration uses a thin corrector plate, which is a shell concentric about the aperture stop. The aperture stop is at the center of curvature of the primary mirror as with the Schmidt telescope. The net result is a level of performance that comes close to the Schmidt telescope but without aspheric surfaces. If an aspheric corrector plate is now located at the aperture stop the aberration correction becomes incredible. This is because the shell corrector is concentric about the stop and only the weak aspheric corrector suffers from obliquity effects.

HYBRID A hybrid system is a combination of several pure or classical solutions. What we have here is a combination of the following:

- A multielement corrector group of a diameter equal to the entrance pupil. This near zero power group of elements is typically three to five lens elements and can be of the same glass type with no chromatic aberrations because it is of zero net optical power. The primary purpose of the corrector group is to balance and cancel the spherical aberration of the spherical primary mirror.
- The bulk of the optical power is from the primary mirror as well as the secondary, which itself is an aluminized area on the aft surface of the corrector group.
- Finally, there is a field-correcting group just before the image plane. This group of elements can effectively flatten the field of view and correct the residual off-axis aberrations such as coma and astigmatism.

The beauty of working with the hybrid system configuration described here is that each functional attribute of the system is quite independent and can be easily understood. The front corrector group corrects the spherical aberration of the primary mirror. The fact that it is of near zero power means that a single glass will produce a result essentially free

of chromatic aberration. And the rear field-correcting group can easily flatten the field curvature and simultaneously correct any residual off-axis aberrations, including coma and astigmatism.

UNOBSCURED APERTURE SYSTEMS Finally, there is a class of reflective optical systems generally known as “unobscured aperture systems.” These include the three-mirror anastigmat (TMA) as well as other forms of all reflective nonrotationally symmetric optical systems. These systems eliminate the performance degrading and difficult to support secondary mirror of the Cassegrain. However, we sometimes require nonrotationally symmetric aspheric surfaces which are difficult to manufacture and test and which are generally costly. Figure 8.6 shows a typical form of TMA. This configuration is an afocal telescope and is per US patent 5,173,801 by Cook. While the most obvious advantage of a TMA is in the hardware, especially as it relates to not needing to support a secondary mirror as in a Cassegrain form of system, the TMA can have significant advantages with respect to the suppression of unwanted diffracted light. Consider Fig. 8.7 where we show a system form developed for space applications. While there are variations on the basic theme, the essence of these systems is that they are used to image relatively close to a bright source as might be encountered in a space application when we are looking within several degrees of the sun. The sky is black, yet the intense solar radiation is just outside our field of view. In order to suppress the diffracted light, the aperture stop is reimaged to a location

Figure 8.6
Three-Mirror
Anastigmat

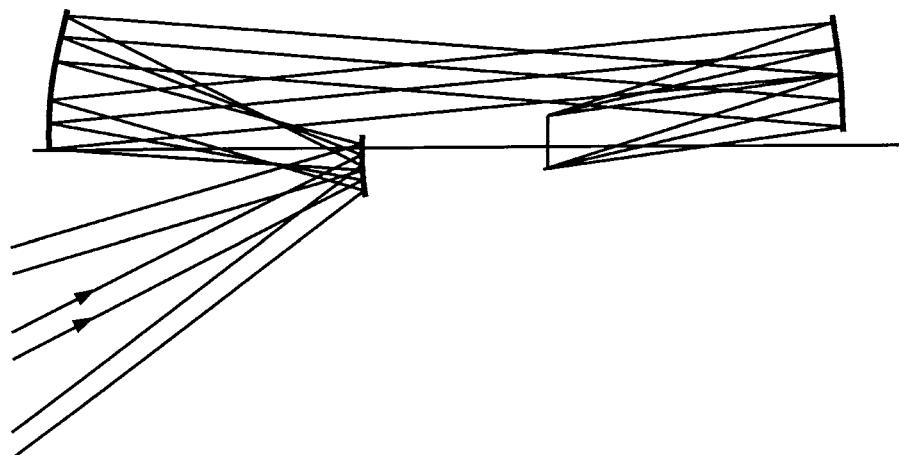
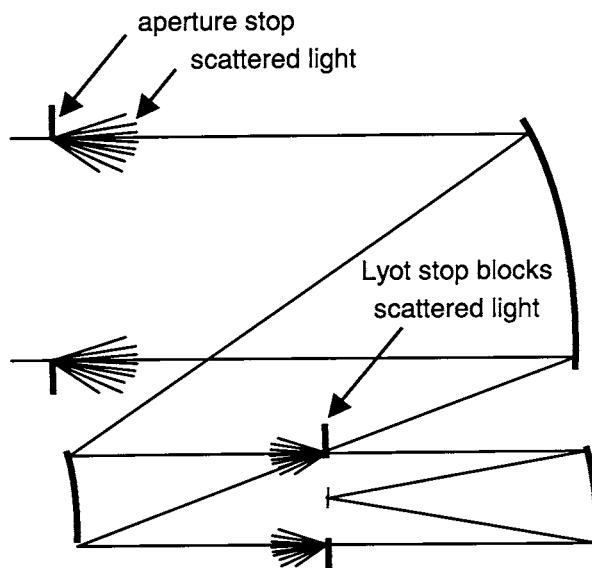


Figure 8.7

Three-Mirror All-Reflective System
Showing Lyot Stop
for Stray Light
Suppression



within the system where an aperture stop slightly smaller than that of the reimaged stop can be located so as to block the reimaged scattered light. This is known as a ‘Lyot stop’ (pronounced “Leo,” after the French astronomer Bernard Lyot).

Reflective Systems, Relative Merits

- *No chromatic aberrations.* There are no chromatic aberrations whatsoever in all-reflecting systems. According to ray tracing theory and the use of Snell’s law, the refractive index of a mirror is -1.0 for all wavelengths. For this reason reflective optics can be extremely well suited for multispectral applications or situations where refractive materials are either expensive or unavailable.
- *Central obscuration.* Since mirrors get in each other’s way, there is often a central obscuration associated with reflective optical systems such as in the Cassegrain configuration. This obscuration affects the net photon throughput, affects the image contrast or MTF (Chap. 15), and is difficult to mount and align. Needless to say,

the support structure for this “secondary mirror” must be of a minimal obscuration to the incoming light, yet it must be strong and robust.

- *Aspheric surfaces required.* Due to the limited number of surfaces that can be effectively used in a reflective optical system, there are rarely enough surfaces to allow for minimization of aberrations as with refractive lens systems. For example, a typical double Gauss lens system may consist of seven elements. If there were no cemented elements, we would have 14 radii with which to minimize ray bendings and hence minimize the residual aberrations. In our Cassegrain reflective system configuration, we have only two surfaces, which are far too few for aberration control, thus leading us to require the use of nonspherical surfaces. One has only to ask the question of how a reflective system with 14 mirrors would look to appreciate the difficulty of working effectively with more than two to three mirrors. It is important to realize that aspheric surfaces are not necessarily bad. There are many contemporary methodologies for producing aspheric surfaces on both mirror as well as transmissive lens surfaces. For mirrors, we have ultraprecision machining or diamond turning as the most common.
- *Small number of elements.* The small number of elements together with the baffling issues, and the fact that the mirrors get in each other’s way, limits the field of view of reflective systems to be generally smaller than of the refractive systems.
- *Can be low weight.* Reflective systems can, in many cases, be made of aluminum, which is light in weight.
- *Inherently athermalized.* Reflective systems, if manufactured of a single material such as aluminum, are generally athermal. In other words, for a uniform temperature increase or decrease, the entire system expands or contracts by an amount dependent on the thermal coefficient of expansion. Since this is a uniform scaling of all system parameters, the image will still be in focus. If multiple materials and/or thermal gradients are used, then a careful assessment of the thermal properties of the imagery is critical. Zerodur is a glass material with almost zero coefficient of thermal expansion, and it is commonly used for large glass mirrors. The important message here is that reflective systems have the potential for being fully athermalized. To convince yourself that all reflective systems manufactured of the same material

throughout are indeed athermal, consider the following explanation: Assume that our design is a Cassegrain reflective system where all components, including the mirrors and the support structure, are aluminum. If we heat or cool the system uniformly, then it will uniformly expand or contract according to the thermal coefficient of expansion of the material. This is, in effect, a scaling of the entire system. Imagine a drawing of our reflective system forming a perfectly focused image. Take this drawing to a copy machine and enlarge it by 20%. Now look at the drawing and ask yourself "is it still in focus?" Of course it is! Needless to say, if you have thermal gradients and/or different materials, your system may not be sufficiently athermal, and this may require active or passive athermalization.

- *Stray light susceptibility.* Reflective systems are often faced with problems associated with stray light. This stray light is often out-of-field light which directly or indirectly reaches the final sensor. The Cassegrain is an example of a system, which needs to be properly baffled to suppress unwanted stray light that may directly go through, missing both mirrors.

Refractive Systems, Relative Merits

- *Straight through.* The system operates straight through without any central obscuration. This results in a potentially higher photon efficiency with none of the degradations associated with a central obscuration.
- *Spherical surfaces, conventional manufacturing.* Since we can add lens elements and use the necessary techniques to minimize aberrations, we can most often use spherical surfaces and thus avoid expensive manufacturing methods often associated with aspheric surfaces.
- *Can add a lot of components.* This makes it possible to design high-speed systems with large fields of view.
- *Expensive materials and athermalization problems in the thermal infrared.* Refractive systems used in the thermal infrared (the MWIR which is the 3- to 5- μm spectral band and the LWIR which is the 8- to 14- μm spectral band) often require materials that are very expensive and have a high dn/dt . For example, germanium is extremely expensive, and furthermore it has a $dn/dt = 0.000396/\text{^{\circ}C}$.

Mirrors and Prisms

Mirrors and prisms are the optical components used in optical systems to

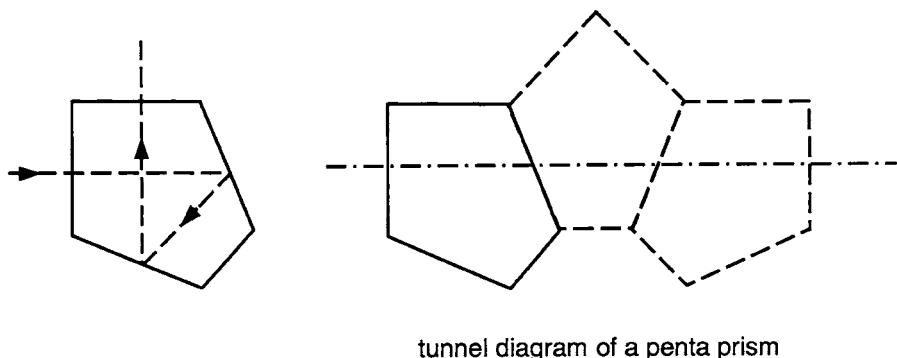
- Change the direction of light.
- Fold an optical system for better packaging.
- Provide a proper image orientation.
- Combine or split the optical beams using beamsplitter coatings.
- Disperse light with wedged prisms.
- Provide the means for interpupillary distance change in binocular systems.
- Expand or contract a laser's beam diameter, etc.

Flat mirrors fold the optical path in a system. Prisms also fold the optical path, except that the reflecting surfaces of the prisms behave like mirrors rigidly mounted with respect to each other. The optical designer has to be careful to leave enough space during the design of the optical system to place the mirrors and prisms where they are needed. Prisms have flat polished surfaces and have no optical power. If prisms are used for the proper image orientation and location, they use refraction at the input and output surfaces and reflection on the intermediate surfaces. Reflection is a total internal reflection if the incident cone of light is small enough and/or the magnitude of the prism angle is such that the condition for the total internal reflection is satisfied for all rays. Otherwise, surfaces are mirror coated.

Reflecting prisms are generally designed so that the entering and exit faces are parallel and perpendicular to the optical axis. This means that the prism can be represented as a plane parallel glass plate. The thickness of the glass plate is obtained by unfolding the prism around its reflecting surfaces. Unfolded prisms are shown in the form of a "tunnel diagram." The Penta prism with its tunnel diagram is shown in Fig. 8.8. We use a tunnel diagram so as to be able to use a single block of glass with no folding of the optical path during the design. The prism of refractive index, n , and the glass thickness, d (thickness of the unfolded prism), has its equivalent path length in air, also called the prism *apparent thickness*, and is equal to

$$d_{\text{apparent}} = \frac{d}{n}$$

Figure 8.8
Penta Prism



tunnel diagram of a penta prism

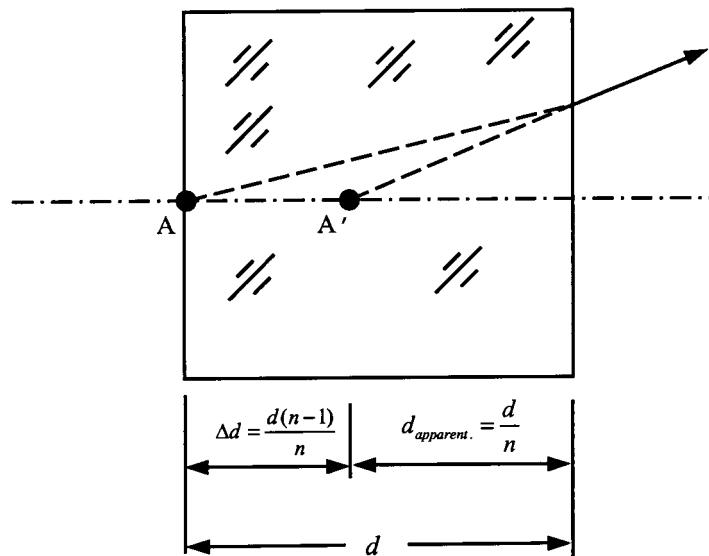
If the prism is inserted in a convergent beam, the image will be shifted by

$$\Delta d = \frac{d(n - 1)}{n}$$

Prism thickness and the apparent prism thickness are shown in Fig. 8.9.

In the case of mirrors, the optical designer has to leave sufficient space in the optical path for the mirrors to be mounted. Prisms do not introduce aberrations only if they are located in a collimated beam.

Figure 8.9
Prism Thickness and
the Apparent Prism
Thickness



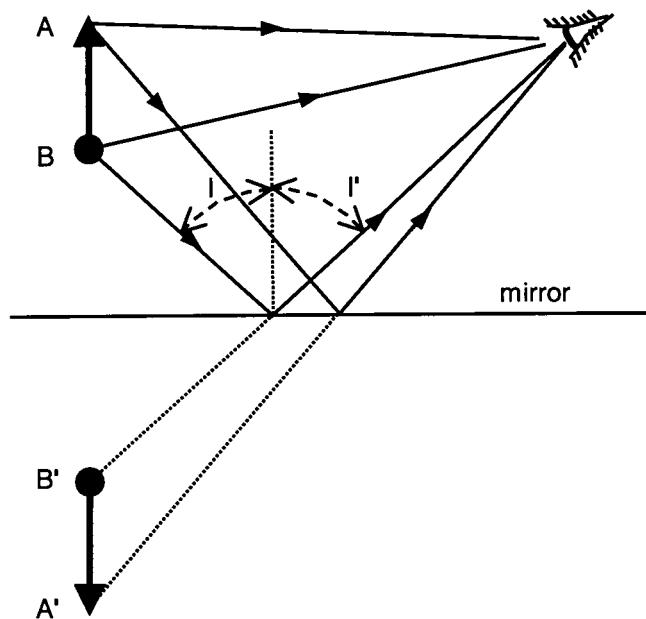
When the optical designer works on a system which contains, among other powered components, a prism of thickness d , he or she usually designs the system using a block of glass of thickness d because of the simplicity, smaller number of surfaces, and, consequently, faster ray tracing. In the case when the prism is located in a convergent or divergent beam, this block of glass introduces both monochromatic and chromatic aberrations, and has to be present during the optimization of the system. The whole process is iterative. The optical designer starts with the rough size of the prism, and optimizes the system with the prism equivalent block of glass. At some point in the design, the designer checks the diameter of the input and the output beam, as well as the ray angles through the prism, and makes the necessary adjustments to the prism size and positioning.

In a real system, the length of the prism along the optical axis is often significantly shorter than the glass equivalent plate, since the optical path in the prism is folded a few times. A good practice is to enter the real prism surfaces in the optical prescription when the design of the system is nearly finished and check the ray footprints on each prism surface as well as the optical performance of the whole system. It is also convenient to export a computer file with all the optical components in the system in a format that is suitable for the mechanical designer to design the mechanics around the optical components.

When a ray is reflected from a flat mirror, the incident ray, the normal to the mirror at the point of incidence, and the reflected ray, all lie in a single plane, which is called the *plane of incidence*.

Let us examine the orientation of the image after reflection from a flat mirror. If the observer looks directly at the object AB shown in Fig. 8.10, the point A appears to be the highest point of the object. If you look at the same object reflected from the mirror, it appears that the image is located behind the mirror in $A'B'$. However, this time it appears that point A' is at the bottom of the image. One reflection changes the orientation of the image in the plane of incidence, which means in our example that the image is upside down. If the object is a two-dimensional object, there is no change in the image orientation in the plane perpendicular to the plane of incidence. In our example, this means that the left side of the object after reflection is seen as coming from the left side. Let us imagine that object AB has no symmetry. It could be, for example, letter R. When viewed directly, the letter is oriented properly, and we can read it. After reflection in the mirror, letter R is upside down, and even if we rotate it around the direction of image

Figure 8.10
Image Orientation
after a Single Reflec-
tion from a Flat Mirror



propagation, we can never orient it so that it will be readable. In the case when the image is readable, which is shown in Fig. 8.11a, we call it a “right-handed image.” If the image is as shown in Fig. 8.11b, where letter R is backwards, regardless of the orientation of the image, it is called a “left-handed image.”

After multiple reflections off flat mirrors, or reflections inside a prism, a general rule says that an even number of reflections gives a right-handed image and an odd number of reflections gives a left-handed image. However, whether the number of reflections is even or odd does not tell us anything about the image orientation.

Figure 8.11
Right- and Left-
Handed Images



(a) Right-handed image



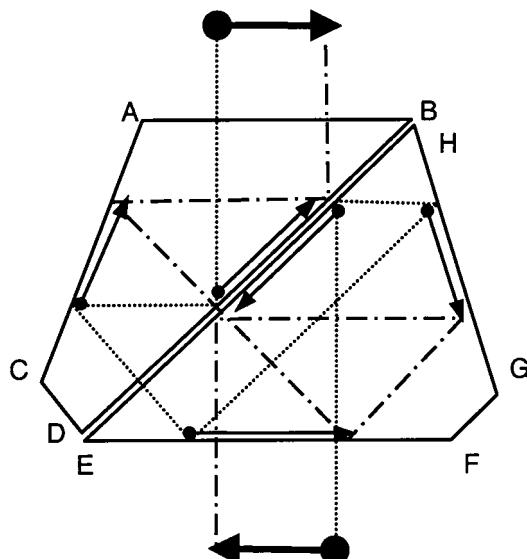
(b) Left-handed image

Reflecting prisms are used to:

- Erect the image in telescopes, which means that the top to bottom as well as the left to right are inverted, and the image is right-handed.
- Invert the image in one plane, either top to bottom or left to right. The image is left-handed.
- Deviate the optical axis, with inversion, erection, or no change in the image orientation.
- Displace the optical axis, with inversion, erection, or no change in the image orientation.
- Keep the unchanged image orientation, rotating the prism around the optical axis (these are prisms with no axis deviation or displacement) while the input image rotates around the center of the field of view (or around the input optical axis).

As an example of an inverting prism, we will look at the *Pechan prism*. A Pechan prism is shown in Fig. 8.12. It consists of two prisms with a small air gap between them. There are a total of five reflections, which means that a right-handed image entering the prism is changed to a left-handed image after the prism. All five reflections have a common plane

Figure 8.12
Pechan Prism



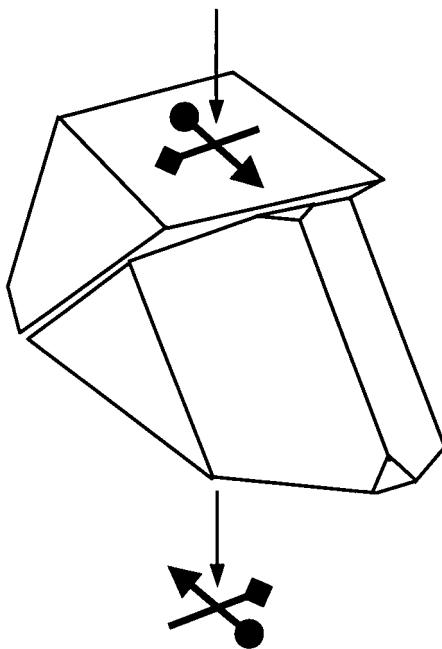
of incidence, and the image is inverted in that plane. In the direction which is normal to the common plane of incidence, there is no change in the image orientation.

Tracing of an image through a Pechan prism is shown in Fig. 8.12. The object chosen for tracing is a circle-arrow object. After refraction through the surface, *AB*, the image is totally internally reflected from surface *BD*. This TIR limits the field of view of the Pechan prism, which is about 10° for medium-index glasses. After surface *BD*, the rays reach the surface, *AC*, which has to be mirror coated. Small arrows on each surface show the orientation of the circle-arrow object as it falls on a given surface. The image is then refracted by the two surfaces with the air gap between them, and then totally internally reflected off surface *EF*. The next surface, *GH*, also has to be mirror coated for the rays to be reflected, and after the final TIR on surface *EH*, the image exits the prism as an inverted image with no deviation. Although the Pechan prism consists of two prisms, which means that the alignment and mounting of the prism is rather complicated, it is very commonly used to invert the image. Most of the systems that require image rotation or derotation, have a Pechan prism that rotates half the rotation angle of the scanning entrance mirror. It can be used in a convergent beam, and the clearance for the prism rotation is the smallest of all inverting prisms with no deviation.

If surface *GH* of the Pechan prism is converted into a roof, where two sides of the roof are normal to each other, the prism is called a roof Pechan prism (Fig. 8.13). It has six reflections, so that the image is right-handed. The added reflection on the roof is in the plane which is normal to all other planes of reflection in the prism. This means that the image is right side up. Binoculars with a straight axis, which are Newton-type telescopes, use this type of prism for image erection. The prism is located between the objective and the image plane formed in the focus of the eyepiece. The light beam coming from one point in the field is split on the roof in two beam segments, each undergoing reflections on both sides of the roof. These two segments join after the roof, and form one image. However, if there is an error in the 90° roof angle, it may introduce the image blur or a double image. The tolerance on a roof angle is, in most cases, only a few arc-seconds.

An optical designer has to be careful when designing the prisms, or using standard prisms in an optical system, because the prisms have to work properly for a specified field of view, or a given cone angle. Mistakes can result in the appearance of ghost images. For example, in the case of the Pechan prism, there are three surfaces where the beam is

Figure 8.13
Roof Pechan Prism



totally internally reflected. If the incident cone angle is too large for a chosen index of refraction of the prism, there will be one part of the field, which will not be properly reflected through the prism as the rest of the image. It will only be refracted and pass directly through the prism. The other case when the ghost images can appear is when the prism is not large enough and some skew rays undergo one extra reflection off the side surfaces. This can happen in the right-angle prism. The ghost reflections may be eliminated making the prism larger, and cutting notches in the prism.

Let us now look at a right-angle telescope, determine what kind of prism can be used to deviate the optical axis by 90° , and also erect the image so that the viewer sees a noninverted right-handed image through the eyepiece. A Newtonian telescope creates an inverted right-handed image. If we put a screen in the focal plane of the objective, the image is both inverted and reverted. The objective creates an image which is right-handed, with an altered orientation both left to right and top to bottom. The eyepiece acts like a magnifier and it does not change the image orientation.

In order to determine the image orientation at any location along the optical system, it is convenient to write on a small piece of paper an object with no symmetry. This can be a letter R, and this piece of paper should be moved through the space, simulating reflections off the prism mirror surfaces. You should always be located such that the image is moving toward your eyes. Do not forget to rotate the image left to right and top to bottom in the case of the objective lens or a relay lens, and do not rotate it through the eyepiece.

Our problem is sketched in Fig. 8.14. The object is shown as a circle at the bottom, an arrow at the top and a square on the left. After the objective, the image changes its orientation in both directions. Since the eyepiece does not change the image orientation, corrections in the image orientation in the horizontal and the vertical plane have to be done with the prism. If we would have only a plane mirror, the arrow and the circle would be properly oriented after the reflection. This is shown with the dotted lines drawn parallel to the optical axis. However, the square, which was turned to the right side after the objective, would stay on the right side after the reflection off the mirror. This means that we have to find a way to invert the image left to right, keeping only one reflection in the vertical plane. This can be accomplished using the right-angle prism and adding the roof on the hypotenuse surface. This prism is called the *Amici prism*, and it is shown in Fig. 8.15.

Figure 8.14
Right-Angle Telescope

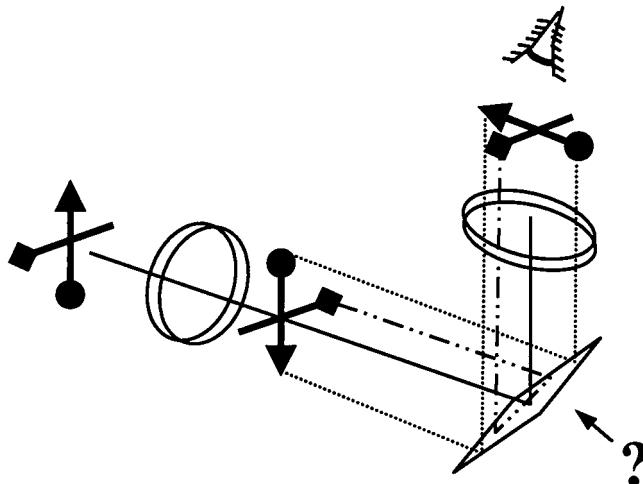
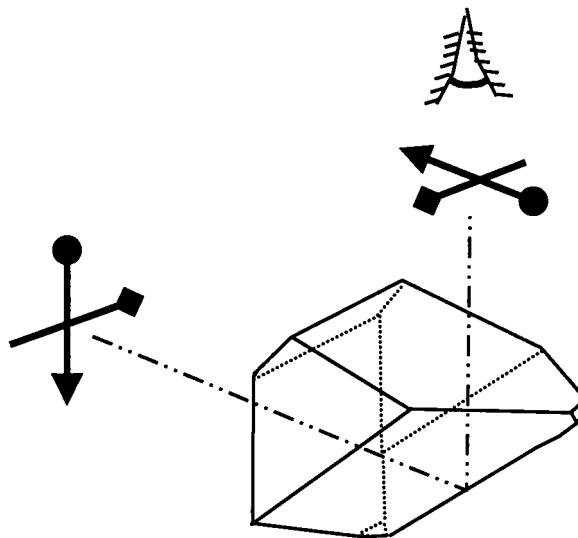


Figure 8.15
Amici Prism



There are many standard prism types. Useful information about prisms, their function, and the important dimensional relationships, can be found in the MIL-HDBK-141 (1962).

Design of Visual Systems

Visual optics includes the wide variety of optical systems creating imagery to be viewed directly by the human eye. This includes telescopes, microscopes, binoculars, riflescopes, camera viewfinders, head-mounted displays, magnifiers, and others. Common to all of these systems is that the eye is looking into some form of viewing optics such as an eyepiece.

Basic Parameters of the Human Eye

The basic optical parameters of the human eye are listed in Table 8.1. Note that most of the data are listed as “approximately” due to the natural variation from person to person.

TABLE 8.1

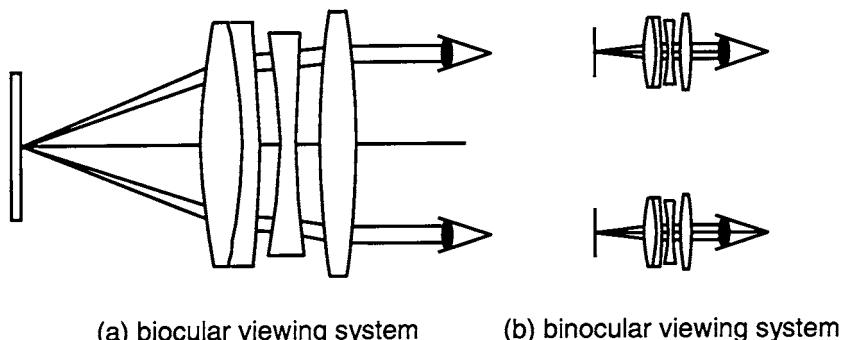
Typical Human Eye
Optical Parameters

Parameter	Value for Human Eye
Entrance pupil diameter (mm)	≈ 2.5 to ≈ 7
Focal length (mm)	≈ 16.9
f/number	≈ 2.4 to ≈ 6.8
Distance from cornea to point of rotation (mm)	≈ 13.5
Radius of cornea (mm)	≈ 8
Interpupillary distance (IPD) (mm)	≈ 63.5 average, 46 to 80 range
Accurate seeing area (degrees)	≈ 1
Normal viewing angles (horizontal) (degrees)	$\approx \pm 5$ to ± 30
Total visual limit (horizontal) (degrees)	$\approx \pm 108$

The eye of an average young person can accommodate or focus to a distance in the order of 250 mm. Due to geometry, as a person looks at a progressively closer object, the two eyes must progressively increase their angle of convergence. In the limit, for an interpupillary distance or IPD of 63.5 mm and an object at a 250-mm distance, the full convergence angle is 14.5° . If we work with a monocular system such as a rifle scope, then convergence is not an issue; however, for a binocular, or even more for biocular systems, it can be quite important. A *biocular* system is a large field-of-view “eyepiece” about 100 mm in diameter, used for viewing a screen or display with both eyes. Figure 8.16 shows a binocular and a biocular system. The design of biocular systems is extremely difficult, and complex multielement designs often are required. One of the difficulties to the designer is that each eye is looking through the extreme edges of the biocular system, and this is the region of the pupil which generally is the most difficult to correct for aberrations.

The user of a pair of binoculars will manually focus the binoculars. Since there is generally no convergence in the optical paths, the user will most often focus the binoculars at infinity, especially when looking at distant objects. This means that collimated light will exit the eyepiece and enter into the eye. If we were to design a rifle scope, convergence is not an issue, so the user will focus the device to the most comfortable

Figure 8.16
Biocular and Binocular Systems



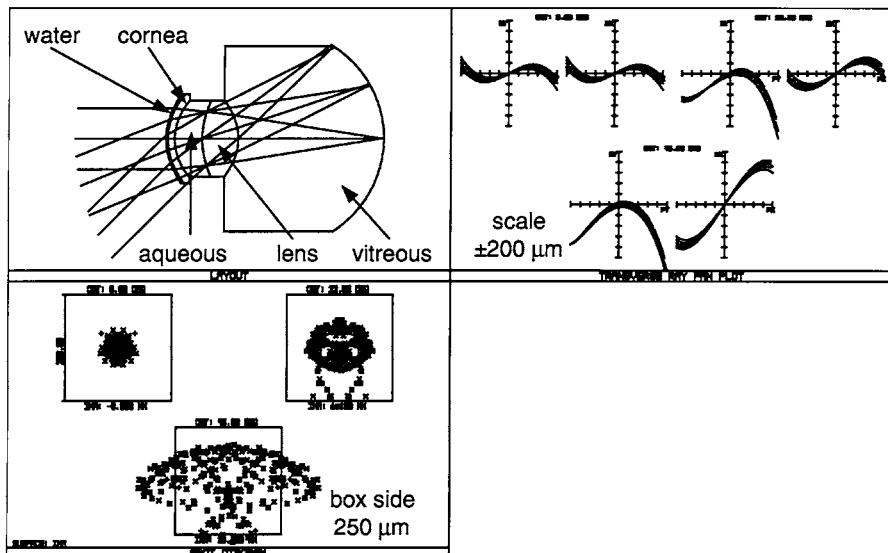
distance. This is often in the range of 2.5 to 6 m, although some research has shown that the resting state of the eye in a dark condition is more like 1 m. It is important to keep in mind that if you require the user to accommodate to some close distance, then you should consider converging the two optical paths (diverging them into the eyes).

You can appreciate the situation if you consider providing diverging light into the two eyes from a virtual object 250 mm in front of the user. While the user can easily accommodate to a 250-mm distance, if the two optical paths were parallel to each other, then there would be potentially significant eye strain, since when accommodating to 250 mm, a person will naturally converge his or her eyes by about 14.5° as noted previously.

There are a number of reasonably reliable eye models, and Figs. 8.17 and 8.18 show one of the more common models, the Lotmar eye model, for eye pupils of 7 and 3 mm in diameter, respectively. The data are shown for fields of view of 0° , 22.5° , and 45° off axis. You can see that there is a residual of spherical aberration which is, of course, more prominent with the larger pupil diameter. For the 7-mm-diameter pupil the spherical aberration equates to an rms blur diameter of about 8 min of arc on the retina, and this reduces to about 1.3 min of arc for the 3-mm-diameter pupil. The eye was permitted to refocus for each pupil diameter. Note the significant off-axis coma and astigmatism residuals.

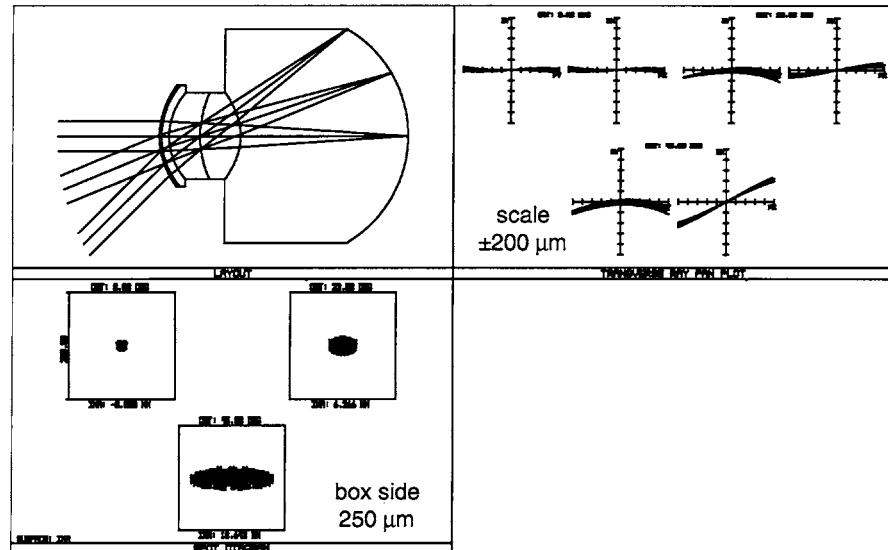
It has been our experience that the residual eye aberrations are somewhat different for different eye models. This leads to the following question: Should the design of a visual system include the effects of the eye? In other words, should your optical design attempt to cancel the eye's residual aberrations based on one of the eye models? In all likelihood

Figure 8.17
Lotmar Eye Model
with 7-mm-Diameter
Pupil



you could do a reasonably good job of accomplishing this from a lens design standpoint; however, the disadvantage is that the eye's aberrations vary from person to person, and even the eye models available give somewhat different aberration residuals. It is generally accepted that it is most prudent to design visual optical systems assuming a perfect eye.

Figure 8.18
Lotmar Eye Model
with 3-mm-Diameter
Pupil

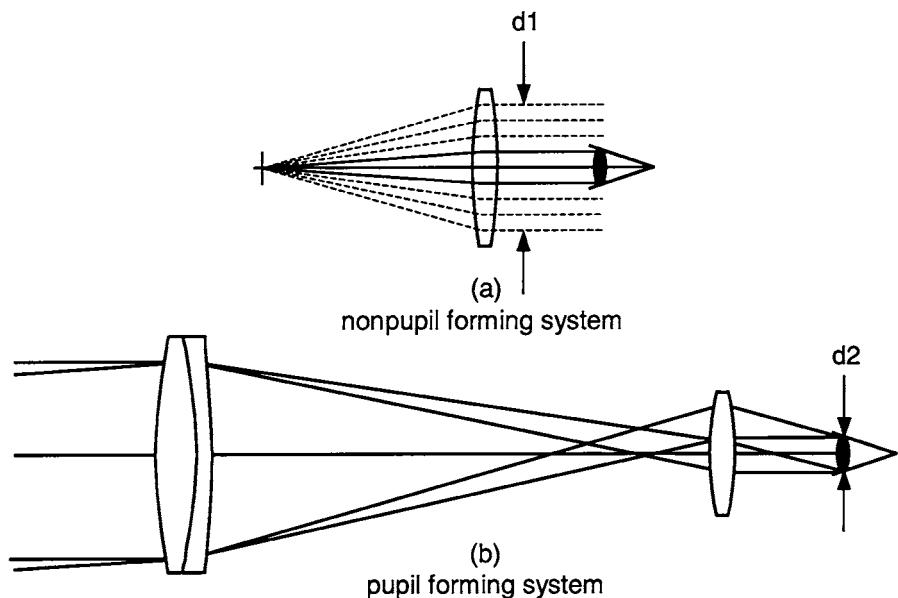


Pupil-Forming Systems

Most (but not all) visual systems are known as “pupil-forming systems.” The term can be best illustrated by considering Fig. 8.19, where we show a magnifying glass in Fig. 8.19a and a simple telescope in Fig. 8.19b. The magnifying glass is *not* a pupil-forming system. What this means is that there is no well-defined exit pupil from the magnifying glass which needs to be mated or matched to the entrance pupil of the eye. On the other hand, the exit pupil of the telescope (Fig. 8.19b) is well defined, and in order to see the imagery, the entrance pupil of the eye *must* be lined up with the exit pupil of the telescope.

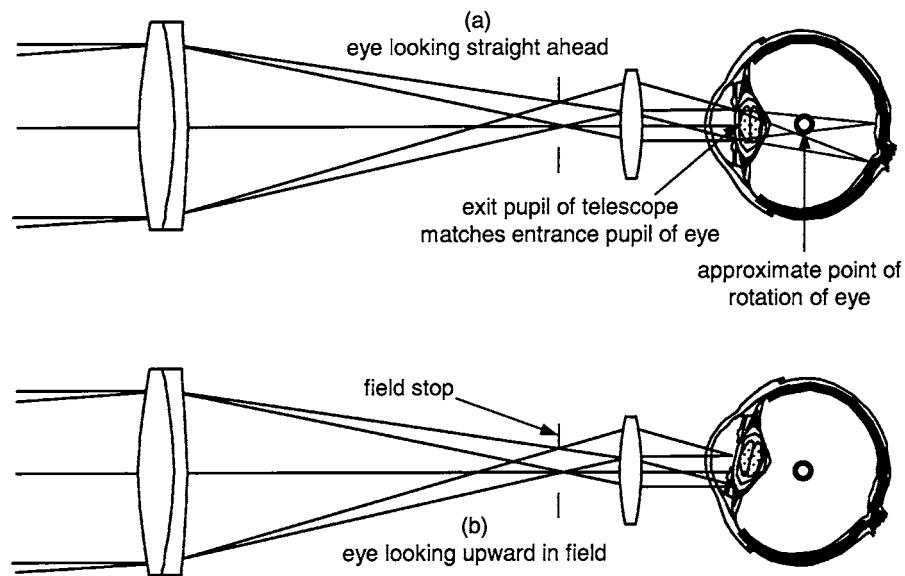
Another way to understand pupil-forming systems is to imagine placing a white card in place of the eye in both Figs. 8.19a and b. Further, assume a bright object which is being viewed. In the case of the magnifying glass, light from the object being viewed will fill a large area of diameter d_1 , which is effectively the diameter of the lens. In the case of the telescope, there will be a bright disk of diameter d_2 at the exit pupil of the telescope. If the telescope were similar to a pair of 7 × 50 binoculars, this bright disk at d_2 will be 7 mm in diameter. The important aspect of pupil-forming systems is that special attention must be given to assure that the light exiting the optical system does enter the pupil of the eye.

Figure 8.19
Pupil-Forming Optical Systems



There is a very interesting subtlety, and this can be understood by looking at Fig. 8.20. In Fig. 8.20a the eye is looking straight ahead to the left. The center of the field of view will be at or near the fovea, which is the highest acuity of the retina, and the user will see in his or her peripheral vision the field stop of the eyepiece. In effect, the imagery will fall within a well-defined circle, and there is black outside of the circle. If the person now looks upward to the left by rotating his or her eye about the center of rotation of the eye, some of the light from the edge of the field will not enter the eye's pupil. In situations where we have a wide apparent field-of-view telescope with a 7-mm or smaller exit pupil diameter, when the person looks upward toward the edge of the field stop, the light may even disappear completely! It is possible to move your eye closer to the eyepiece so that the exit pupil of the telescope is located at the point of rotation of the eye. Now the person can see clearly the field stop in the eyepiece when looking toward the field stop; however, when looking straight ahead, the field stop and the outer periphery of the field of view may completely disappear from view! This is a very striking, as well as a weird, effect, and if you ever have the opportunity to see it, it is worthwhile to do so.

Figure 8.20
Telescope Showing
Effect of Eye Rotation



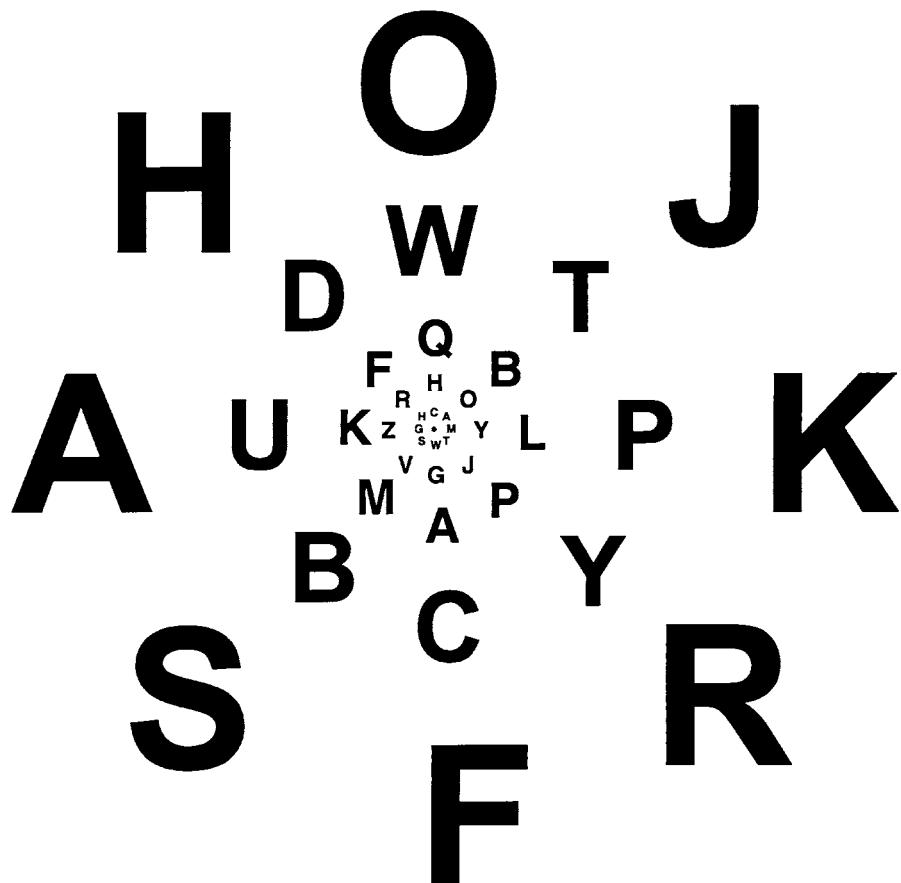
Requirements for Visual Optical Systems

When designing visual systems, a number of parameters must be considered which are unique to these systems, and these are listed here:

- For many hand-held systems such as binoculars, the user will most often bring the object of greatest interest to the center of the field of view. For this reason, off-axis performance is generally more forgiving for these systems.
- The *eye relief* is the clearance or distance from the last optical element in the viewing optics to the front of the cornea of the eye. For eyeglass users the generally accepted minimum eye relief is in the order of 25 mm. Larger values are, of course, helpful, but the diameter of the eyepiece grows with the eye relief.
- It is clear from the transverse ray aberration curves for the Lotmar eye model that the resolution of the eye degrades quite severely as we move away from the center of the field of view. The visual acuity of the eye is maximum at the fovea, which is very close to the center of the field of view of the eye. As an object moves away from the fovea, the visual acuity decreases dramatically and at $\pm 20^\circ$ the visual acuity is only about 10% of that at the fovea. This is well illustrated in Fig. 8.21 where we show a series of letters of different sizes. The increase in letter size is inversely proportional to the decrease in visual acuity as we move from the fovea. Thus, if you look at the small spot at the center of the figure, all of the letters should be approximately equally resolved.

Although it is true that the visual acuity drops quickly away from the fovea, visual systems should have relatively good imagery at the edge of the field. How good depends on the application of the system. For example, consider a binocular. A viewer can rotate his or her eyes to look toward the edge of the field, in which case the field periphery is the sharpest area the eye sees. This could lead us to a conclusion that the edge of the field should be very well corrected for aberrations. However, we all know that any object in the field can be brought to the center by simply rotating the whole binocular. So the quality of the imagery at the edge of the field in a binocular does not have to be as good as in the center, but it should not be so bad that the image blurring and coloring is immediately noticed when the eyes are pointed to the field edge.

Figure 8.21
Illustration of Visual Acuity As a Function of Distance from the Fovea



- The most general and accepted metric that we often hear is that *the eye resolves 1 min of arc*. What this means is that the eye can resolve the capital letter E when each of the dark or bright bars forming the horizontal portions of the letter subtend 1 min of arc, as shown in Fig. 8.22. This means that the eye resolves 2 min per line pair, or 0.5 line pair per minute of arc.
- With respect to visual systems design, providing an image blur diameter from 1 min of arc to perhaps 3 min of arc is generally considered an acceptable level of performance in the center of the field of view. At the edge of the field, 20 to 40 arc min of image blur maximum may be acceptable.

Figure 8.22
Eye Resolution

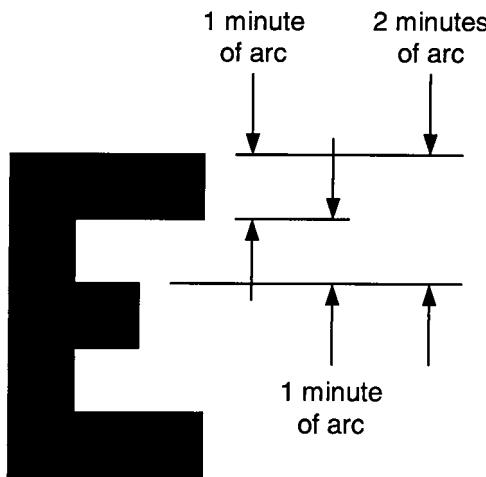


Table 8.2 shows a list of the typical tolerances associated with optics for the eye, as developed by Mouroulis. These data are for binocular viewing with two eyes. Note that items that the eye normally does not do are specified with tighter tolerances. This includes dipvergence (one eye looking upward and the other eye downward) and divergence for example.

There is a rule of thumb that says that the eye can resolve a contrast of 5%. Thus, it is not uncommon to determine where the MTF drops to 0.05 and to conclude that the eye will resolve this spatial frequency. There is perhaps a better way, and this is to use the so-called aerial image modulation (AIM) curve. This is the relationship between the contrast and the number of line pairs per millimeter that the eye can resolve. Walker shows these data, and they are summarized in Fig. 8.23. The AIM curve is a relationship between the modulation required to resolve a given target and the spatial frequency of the target. Using the Lotmar eye model with a 7-mm pupil diameter, we see that if the eye were diffraction limited in Fig. 8.23a, we would resolve 1.6 min/line pair, or 0.8 min/line; however, due to the spherical aberration, we can only resolve in the order of 3 min/line pair, or 1.5 min/line (Fig. 8.23b). In the case of a 3-mm pupil diameter the diffraction-limited eye would resolve about 1.6 mm/line pair (Fig. 8.23c) and the Lotmar model predicts about 1.7 mm/line pair (Fig. 8.23d). It is important to realize that these data are only as accurate as the Lotmar eye model and the referenced AIM threshold data. While they may not be precisely accurate, they do give us a good indication of the resolution of the eye.

TABLE 8.2

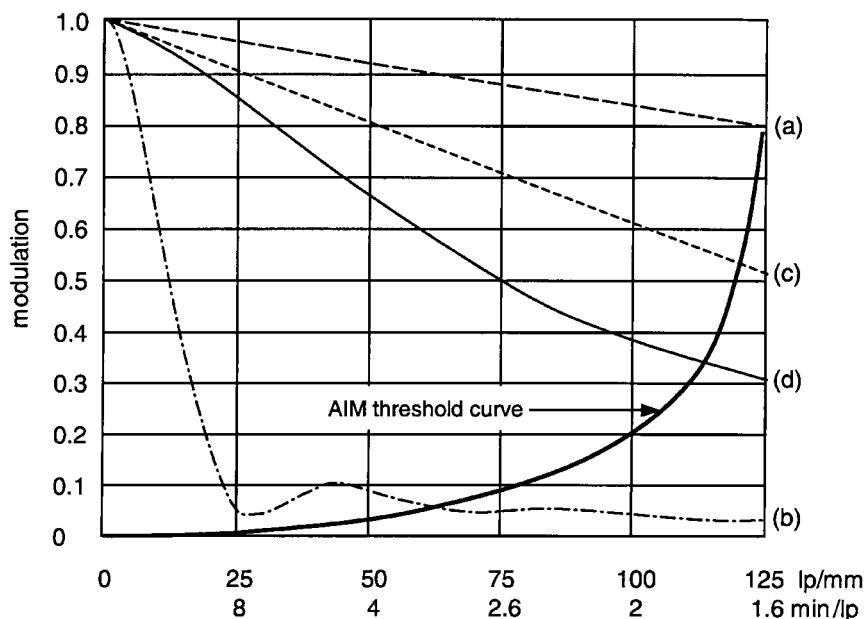
Typical Tolerances
for Visual Systems

Parameter	Typical Specification
Divergence (degrees)	0
Convergence (degrees)	1.5
Dipvergence (min)	8
Magnification difference (%)	± 0.5 to ± 1
Brightness difference (%)	± 10
Best focus if fixed	-1 diopter*
Accommodation	-3 diopter to infinity is OK -1 to -1.5 diopter is best
Axial chromatic aberration	-1 to -1.5 diopters
Lateral color (min)	2
Image quality	Integral of MTF from 0 to 20 line pairs/degree

*A diopter is the reciprocal of focal length, in meters. As used in the table, -1 diopter means that the eye will be focusing to a distance of 1 m in front of the user.

Figure 8.23

Use of AIM Threshold Curve to Predict Visual Resolution



Distortion is an important criterion in visual systems. As we learned in Chap. 5, distortion is more of a mapping error rather than an image-degrading aberration. Its primary effect is to the cosmetic appearance of the image. Distortion is a change in magnification with field of view. Generally, positive or negative distortion in the order of 2 to 2.5% is small enough to be almost imperceptible. Large amounts of distortion can be annoying in any visual system. Telescopes with a large apparent field of view often have up to 10% of distortion at the edge of the field.

Summary on Design of Visual Systems

It is best in the design of visual optical systems to assume a perfect eye model. Most of the computer design programs have the ability to model a so-called perfect lens or paraxial lens, which is used following the system being designed to emulate the eye (or any form of focusing lens). What these so-called perfect lenses do is to simply compute the angular aberrations in the optical system and multiply them by the focal length of the perfect lens which is input by the user. So if you are interested in the actual image blur on the retina, then use a perfect lens of focal length 16.9 mm, which is the approximate focal length of the eye.

If you are designing a visual system which must be fixed focus, it is probably best to design and produce your system to provide light which is either collimated to the user or appears to come from a distance of from 3 to 6 m. Some data suggest a value closer to 1 m. If you can allow the user to refocus, this is generally preferred.

The image quality, or blur diameter, should be in the order of 1 min of arc to 3 min of arc. The specific application will have a lot to do with how good the imagery needs to be. You should design your system for an exit pupil diameter of at least 7 mm (10 to 12 mm is better), and then evaluate the performance both at the 7-mm-diameter pupil as well as with reduced pupil diameters. Some detailed specifications include the eye being decentered to various positions within the exit pupil of the optical system. If the system is to be used in bright conditions, you may even evaluate your design in the 2.5- to 3-mm pupil diameter region.

Our final note on the design of visual systems is to keep in mind that the eye is quite forgiving. Persons who get a new eyeglass prescription

often notice color fringing or lateral color when looking toward the outer periphery of their field of regard (30° to 40° from straight ahead, for example). After several days to several weeks, this color fringing often seems to disappear. The eye, along with the rest of the human visual system, is, in effect, a very powerful computer with impressive image-processing capabilities. This does not mean that you can ignore the aberrations and image quality of visual systems, what it does mean, however, is that, in many situations, the optics for visual systems can be more forgiving than you might think.

The Optical Design Process

The optical design process includes a myriad of tasks that the designer must perform and consider in the process of optimizing the performance of an imaging optical system. While we often think primarily of the robustness of the optimization algorithm, reduction of aberrations, and the like, there is much more to do. The designer must be at what we sometimes call “mental and technical equilibrium with the task at hand.” This means that he or she needs to be fully confident that all of the following are understood and under control:

- All first-order parameters and specifications such as magnification, focal length, *f*/number, full field of view, spectral band and relative weightings, and others.
- Assure that the optical performance is being met, including image quality, distortion, vignetting, and others.
- Assure that the packaging and other physical requirements, including the thermal environment, is being taken into account.
- Assure that the design is manufacturable at a reasonable cost based on a fabrication, assembly, and alignment tolerance analysis and performance error budget.
- Consider all possible problems such as polarization effects, including birefringence, coating feasibility, ghost images and stray light, and any other possible problems.

Once every one of these items has been addressed and is at least recognized and understood, we start with the sketch of the system. First, the system is divided into subsystems if possible, and the first-order parameters are determined for each subsystem. For example, if we are to design a telescope with a given magnification, the entrance pupil diameter should be chosen such that the exit pupil size matches the eye pupil. A focal length of the objective and the eyepiece should be chosen such that the eyepiece can have a sufficiently large eye relief. Now, when the specs for each subsystem are defined, it is time to use the computer-aided design algorithms and associated software to optimize the system, which will be discussed in the rest of this chapter. Each subsystem can be designed and optimized individually and the modules joined together or, more often, some subsystems are optimized separately and some as an integral part of the whole system.

What Do We Do When We Optimize a Lens System?

Present-day computer hardware and software have significantly changed the process of lens design. A simple lens with several elements has nearly an infinite number of possible solutions. Each surface can take on an infinite number of specific radii, ranging from steeply curved concave, through flat, and on to steeply curved convex. There are a near infinite number of possible design permutations for even the simplest lenses. How does one optimize the performance with so many possible permutations? Computers have made what was once a tedious and time-consuming task at least manageable.

The essence of most lens design computer programs is as follows:

- First, the designer has to enter in the program the starting optical system. Then, each variable is changed a small amount, called an *increment*, and the effect to performance is then computed. For example, the first thickness may be changed by 0.05 mm as its increment. Once this increment in thickness is made, the overall performance, including image quality as well as physical constraints, are computed. The results are stored, and the second thickness is now changed by 0.05 mm and so on for all variables that the user has designated. Variables include radii, airspaces,

element thicknesses, glass refractive index, and Abbe number. If you are using aspheric or diffractive surfaces, then the appropriate coefficients are also variables.

- The measure of performance as used here is a quantitative characterization of the optical performance combined with a measure of how well the system meets its first-order constraints set by the user such as focal length, packaging constraints, center and edge thickness violations, and others. The result of the computation is a single number called an *error function* or *merit function*. The lower the number, the better the performance. One typical error function criteria is the rms blur radius, which, in effect, is the radius of a circle containing 68% of the energy. Other criteria include optical path difference, and even MTF as described in Chap. 15.
- The result is a series of derivatives relating the change in performance (P) versus the change in the first variable (V_1), the change in performance (P) versus the change in the second variable (V_2), and so on. This takes on the following form:

$$\frac{\partial P}{\partial V_1}, \frac{\partial P}{\partial V_2}, \frac{\partial P}{\partial V_3}, \frac{\partial P}{\partial V_4} \dots$$

- This set of partial derivatives tells in which direction each parameter has to change to reduce the value of the sum of the squares of the performance residuals. This process of simultaneous parameter changes is repeated until an optimum solution is reached.

A lens system consists of a nearly infinite number of possible solutions in a highly multidimensional space, and it is the job of the designer to determine the optimum solution.

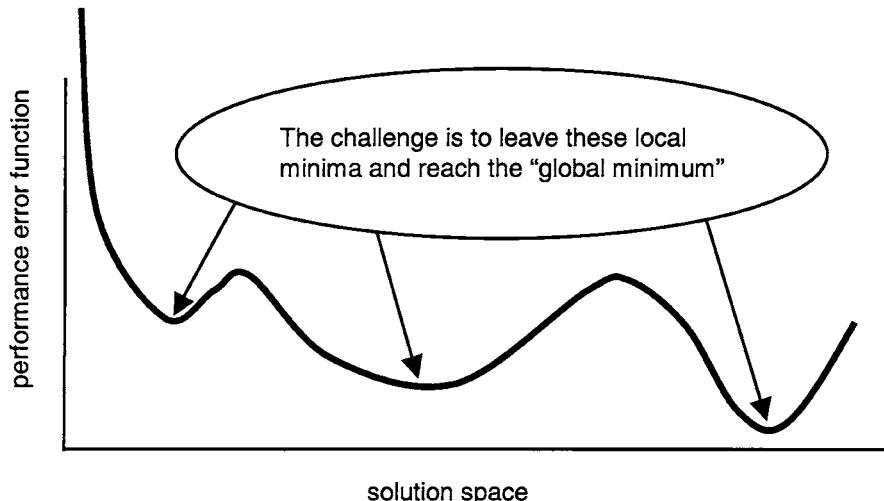
Designers have used the following analogy to describe just how a lens design program works:

- Assume that you cannot see and you are placed in a three-dimensional terrain with randomly changing hills and valleys. Your goal is to locate the lowest elevation or altitude, which in our analogy equates to the lowest error function or merit function. The lower the error function, the better the image quality, with the “goodness” of performance being inversely proportional to the elevation.
- You are given a stick about 2 m long, and you first stand in place and turn around tapping the stick on the ground trying to find which direction to walk so as to go down in elevation.

- Once you determine the azimuth resulting in the greatest drop in elevation, you step forward in that direction by 2 m.
- You now repeat this process until in every direction the elevation goes up or is level, in which case you have located the lowest elevation.
- But what if just over a nearby hill is an even lower valley than you are now in? How can you find this region of solution? You could use a longer stick, or you could step forward a distance several times as long as the length of your stick. If you knew that the derivative or slope downward is linear or at least will continue to proceed downward, this may be a viable approach. This is clearly a nontrivial mathematical problem for which many complex and innovative algorithms have been derived over the years. But the problem is so nontrivial as well as nonlinear that software algorithms to locate the so-called global minimum in the error function are still elusive. Needless to say, the true global minimum in the error function may be quite different or distant from the current location in our n -dimensional terrain.

Figure 9.1 shows a two-dimensional representation of solution space as discussed previously. The *ordinate* is the error function or merit function, which is a measure of image quality, and the *abscissa* is, in effect, solution space. We may initiate a design on the left and the initial

Figure 9.1
Illustration of Solution Space in Lens Design



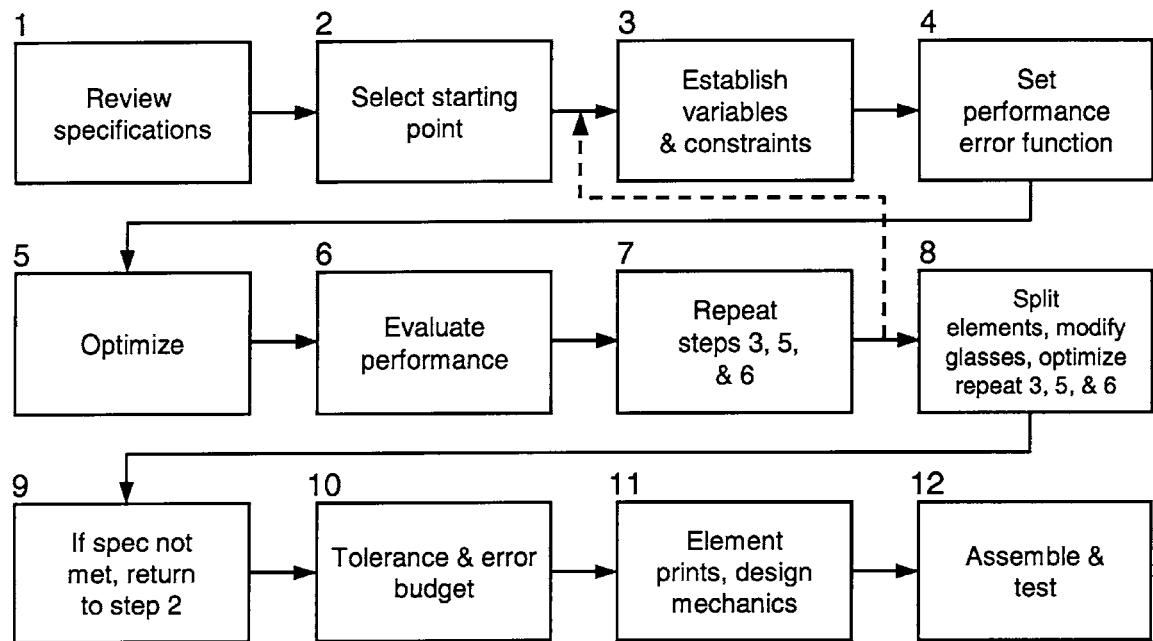
optimization brings the error function to the first minimum called a *local minimum* in the error function. We then change glasses and/or make other changes to the design and ultimately are able to move the design to the next lower local minimum. Finally, we add additional elements and make other changes and we may be able to reach the local minimum on the right. But how do we know that we are at, or even close to, a global minimum? Here lies the challenge as well as the excitement of lens design!

It is important here to note that reaching global minimum in the error function is not necessarily the end goal for a design. Factors including tolerance sensitivity, packaging, viability of materials, number of elements, and many other factors influence the overall assessment or “goodness” of a design. Learning how to optimize a lens system is, of course, quite critical to the overall effort, and learning how to reach a viable local or near-global minimum in the error function is very important to the overall success of a project.

How Does the Designer Approach the Optical Design Task?

The following are the basic steps generally followed by an experienced optical designer in performing a given design task. Needless to say, due to the inherent complexity of optical design, the processes often become far more involved and time consuming. Figure 9.2 outlines these basic steps:

1. The first step in the design process is to *acquire and review all of the specifications*. This includes all optical specifications including focal length, *f*/number, full field of view, packaging constraints, performance goal, environmental requirements, and others.
2. Then we *select a representative viable starting point*. The starting point should, wherever possible, be a configuration which is inherently capable of meeting the specifications for the design. For example, if the specifications are for an *f*/10 monochromatic lens covering a very small field of view and having an entrance pupil diameter of 5 mm, then the lens may very well be a single element. However, if the requirements call for an *f*/1.2 lens over a wide spectral band covering a 40° full field of view, then the solution may very well

**Figure 9.2**

Lens Design and Optimization Procedure

be a very complex six- to seven-element double Gauss lens form. If we were to use a single element for this latter starting point, there would be no hope for a viable solution. Finding a good starting point is very important in obtaining a viable solution. The following are viable sources for starting points:

You can use a *patent* as a starting point. There are many sources for lens patents including Warren Smith's excellent book *Modern Lens Design*. There is also a CD-ROM called "LensView," which contains over 20,000 designs from patents. These are all searchable by a host of key parameters. While the authors of this book are not patent attorneys, we can say with confidence that you may legally enter design data from a patent into your computer and work with it in any way that you would like to. If your resulting design is sold on the market, and if the design infringes on the patent you used (or any other for that matter), you could be cited for patent infringement. It is interesting to note that the purpose

of our patent system in this country is to promote inventions and innovation. This is done by offering an inventor a 17-year exclusive right to his or her invention in exchange for *teaching* in the patent how to implement the invention. Thus, you are, in effect, invited to use the design data and work with it with the goal of coming up with a better design, which you can then go out and patent. By this philosophy, inventors are constantly challenged to improve upon an invention, which, in effect, advances technology, which is what the patent process is all about. Needless to say, we urge you to be careful in your use of patents.

You could use a so-called hybrid design. We mean a hybrid to be the combining of two or more otherwise viable design approaches so as to yield a new system configuration. For example, a moderate field-of-view Tessar lens design form can be combined with one or more strongly negatively powered elements in the front to create an extremely wide-angle lens. In effect, the Tessar is now used over a field of view similar to its designed field, and the negative element or elements bend or "horse" the rays around to cover the wider field of view. An *original design* can, of course, be a viable starting point. As your experience continues to mature, you will eventually become comfortable with "starting from scratch." With today's computer-aided design software, this works most of the time with simple systems such as doublets and triplets; however, with more complex systems, you may have problems and will likely be better off resorting to a patent or other source for a starting point.

3. Once you have entered your starting point into the software package you are using, it is time to *establish the variables and constraints*. The system variables include the following: radii, thicknesses, airspaces, surface tilts and decenters, glass characteristics (refractive index and Abbe number), and aspheric and/or other surface variables, including aspheric coefficients. The constraints include items such as focal length, *f*/number, packaging-related parameters (length, diameter, etc.), specific airspaces, specific ray angles, and virtually any other system requirement. Wavelength and field weights are also required to be input. It is important to note that it is not imperative (nor is it advisable) to vary every conceivable variable in a lens, especially early in the design phase. For example, your initial design optimization should

probably be done using the glasses from the starting point, in other words do not vary glass characteristics initially. This will come later once the design begins to take shape and becomes viable. You may also want to restrict the radii or thicknesses you vary as well, at least initially. For example, if adjacent elements have a very small airspace in the starting design, this may be for a good reason, and you should probably leave them fixed. Also, element thicknesses are very often not of great value as variables, at least initially, in a design task, so it is usually best to keep element thicknesses set to values which will be viable for the manufacturer.

4. You now will *set the performance error function and enter the constraints*. Most programs allow the user to define a fully “canned” or automatically generated error function, which, as discussed earlier, may be the rms blur radius weighted over the input wavelengths and the fields of view. In the Zemax program the user selects the number of rings and arms for which rays will be traced into the entrance pupil (rays are traced at the respective intersection points of the designated number of rings and arms). Chapter 22 shows a detailed example of how we work with the error function.
5. It is now time to *initiate the optimization*. The optimization will run anywhere from a few seconds for simple systems to many hours, depending on just how complex your system is and how many rays, fields of view, wavelengths, and other criteria are in the system. Today, a state-of-the-art PC optimizing a six- to seven-element double Gauss lens with five fields of view will take in the order of 5 to 10 s per optimization cycle. Once the computer has done as much as it can and reaches a local minimum in the error function, it stops and you are automatically exited from the optimization routine.
6. You now *evaluate the performance* using whatever criteria were specified for the lens. This may include MTF, encircled energy, rms spot radius, distortion, and others.
7. You now *repeat steps 3 and 5 until the desired performance is met*. Step 3 was to establish the variables and constraints, and step 5 was to run the optimization, and these steps are repeated as many times as necessary to meet the performance goals. You will often reach a solution that simply does not meet your performance requirements. This is very common during the design evolution, so do not be surprised, depressed, or embarrassed if it happens

to you... it happens to the best of us. When it does happen, you may need to *add or split the optical power of one or more of the lens elements* and/or to *modify glass characteristics*. As we have discussed previously, splitting optical power is extremely valuable in minimizing the aberrations of a lens.

8. There is a really simple way of splitting an element in two, and while it is not “technically robust,” it does work most of the time. What you do is insert two surfaces in the middle of the current element, the first of which will be air and the second is the material of your initial element. The thickness of each “new” element is one-half of the initial element and the airspace should be small, like 0.1, for example. Now simply enter twice the radius of the original element for both s_1 and s_2 of the new elements. You will end up with two elements whose net power sum is nearly the same as your initial element. You can now proceed and vary their radii, the airspace, and, as required, the thicknesses.
9. If you still cannot reach a viable design, then at this point you will need to *return to step 2 and select a new starting point*.
10. Your final task in the design process is to *perform a tolerance analysis and performance error budget*. We will be discussing tolerancing in more depth in Chap. 16. In reality, you should be monitoring your tolerance sensitivities throughout the design process so that if the tolerances appear too tight, you can take action early in the design phase and perhaps select a less sensitive design form.
11. Finally, you will need to *generate optical element prints, contact a viable lens manufacturer, and have your elements produced*. You will also need to work with a qualified mechanical designer who will *design the cell or housing* as well as any required interfaces. It is important to note that while we list the mechanical design as taking place at this point after the lens design is complete, it is extremely important to work with your mechanical designer throughout the lens design process so as to reach an optimum for both the optics as well as the mechanics. Similarly, you should establish a dialog with the optical shop prior to completing the design so as to have time to modify parameters which the shop feels needs attention such as element thicknesses, glass types, and other parameters.
12. Once the components are in house, you will need to *have the lens assembled and tested*. Assembly should be done to a level of precision and cleanliness commensurate with the overall

performance goals. Similarly, testing should be to a criterion which matches or can be correlated with your system specifications and requirements. We discuss testing in Chap. 15.

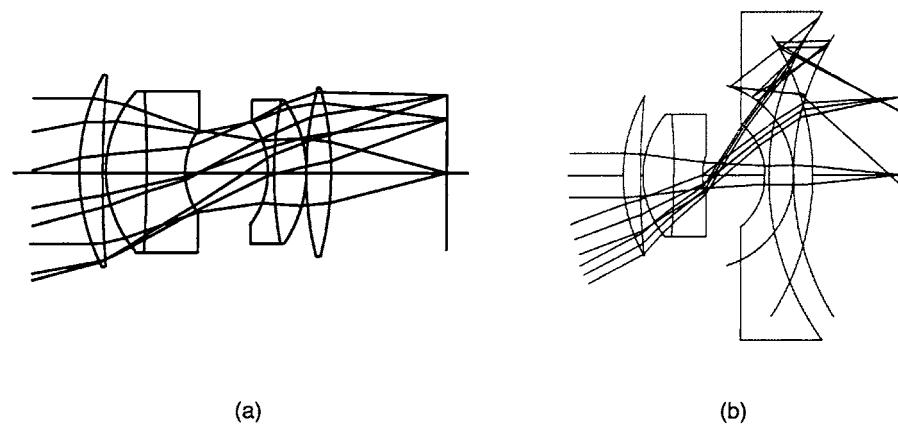
Sample Lens Design Problem

There was a very interesting sample lens design problem presented at the 1980 International Lens Design Conference. The optimized design for an $f/2.0$, 100-mm focal length, 30° full field-of-view double Gauss lens similar to a 35-mm camera lens was sent out to the lens design community. One of the tasks was to redesign the lens to be $f/5$ covering a 55° full field with 50% vignetting permitted. Figure 9.3 shows the original starting design, as well as the design after changing the f /number and field of view, without any optimization.

Sixteen designers submitted their results, and they spent from 2 to 80 h working on the problem. We will present here three representative solutions in Fig. 9.4. The design in Fig. 9.4a is what we often call a *happy lens*. What we mean is that the lens is quite well behaved with no steep bending or severe angles of incidence. The rays seem to “meander” nicely through the lens. It is a comfortable design. We show to the right of the layout a plot of the MTF. MTF will be discussed in detail in Chap. 10. For the purpose of this discussion, consider the MTF to be contrast plotted in the ordinate as a function of the number of line pairs per millimeter

Figure 9.3

Starting Design for Sample Lens Design Problem



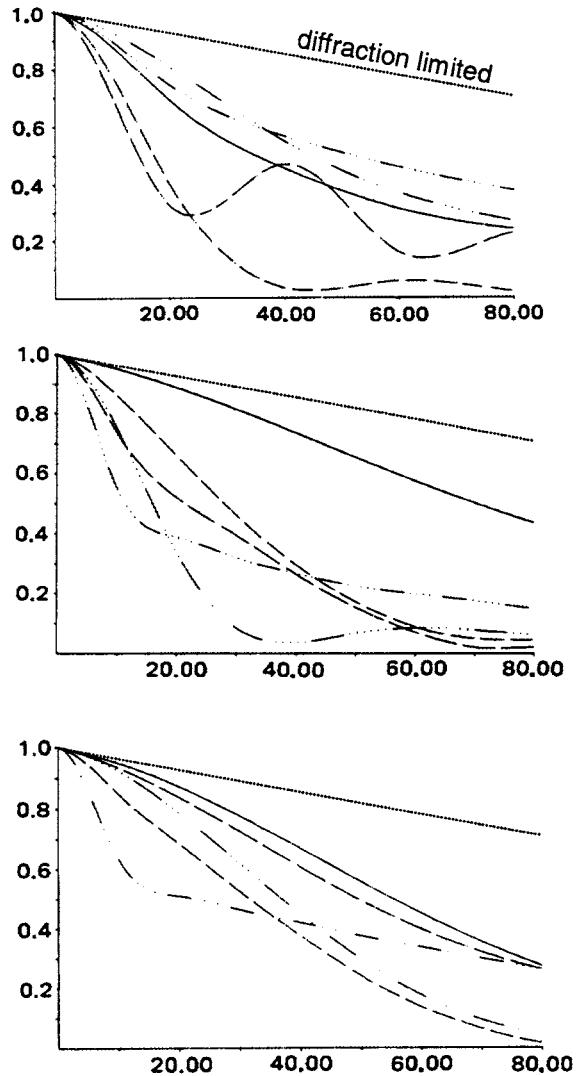
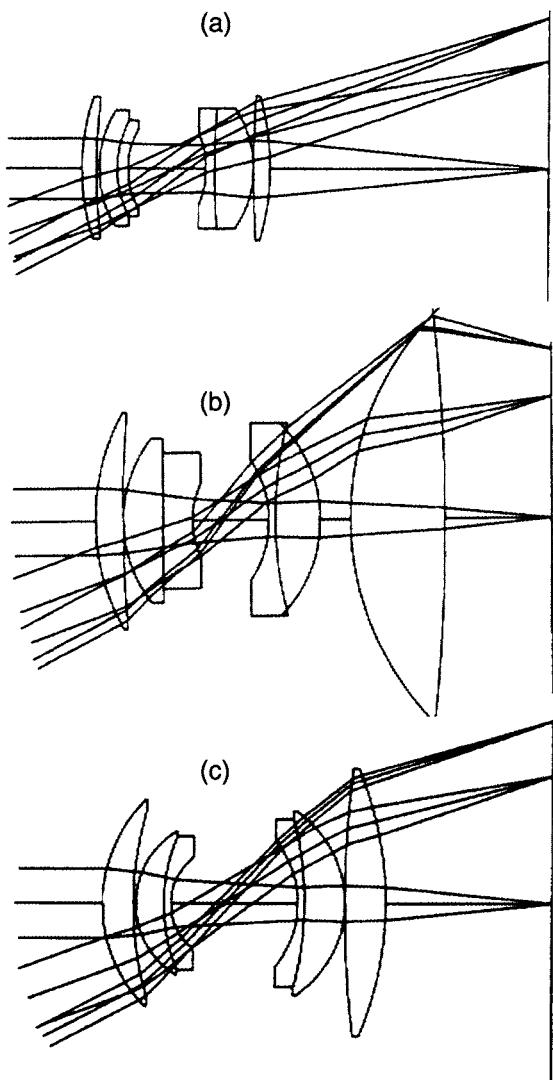


Figure 9.4

Representative Solutions to Sample Lens Design Problem

in the abscissa. The different curves represent different positions in the field of view and different orientations of the resolution patterns. The higher the curves, the better the contrast and the overall performance. The MTF is reasonable for most of the field positions. As will be discussed in Chap. 22, a good rule of thumb for the MTF of a 35-mm camera lens is an MTF of 0.3 at 50 line pairs/mm and 0.5 at 30 line pairs/mm.

The design in Fig. 9.4*b* has a serious problem; the rays entering the last element are at near-grazing angles of incidence. Notice that the exit pupil at full field is to the right of the lens (since the ray cone is descending toward the axis to the right), and at 70% of the field the exit pupil is to the left of the lens (since the ray cone is ascending to the right and therefore appears to have crossed the axis to the left of the lens). This is a direct result of the steep angles of incidence of rays entering the last element. The variation in exit pupil location described here would not itself be an issue unless this lens were used in conjunction with another optical system following it to the right; however, it does indicate clearly the presence of the severe ray bending which will inevitably lead to tight manufacturing and assembly tolerances. Further, the last element has a near-zero edge thickness which would need to be increased. The lens is large, bulky, and heavy. And finally, the MTF of this design is the lowest of the three designs presented.

Finally, the design in Fig. 9.4*c* is somewhat of a compromise of the two prior designs in that it is somewhat spread out from the design in Fig. 9.4*a* but does not have the problems of the design in Fig. 9.4*b*. The MTF of the design in Fig. 9.4*c* is the best of the three designs.

Comparing the three designs is very instructive as it shows the extreme variability of results to the same problem by three designers. The question to ask yourself is what would you do if you subcontracted the design for such a lens, and after a week or two the designer brought you a stack of paper 200-mm thick with the results of the design in Fig. 9.4*b*. And what if he or she said "wow, what a difficult design! But I have this fabulous solution for you!" Prior to reading this book, you might have been inclined to congratulate the designer on a job well done, only to have problems later on during manufacturing and assembly. Now, however, you know that there may be alternate solutions offering superior performance with looser tolerances and improved packaging. Remember that even a simple lens has a near infinite number of possible solutions in a multidimensional space.

Computer Performance Evaluation

What Is Meant by Performance Evaluation

The performance characteristics of an imaging optical system can be represented in many ways. Often the final optical performance specification is in terms of the modulation transfer function (MTF), encircled energy, rms blur diameter, or other image quality criteria. These criteria relate in different ways to the image quality of the system. Image quality can be thought of as *resolution* or how close two objects can approach each other while still being resolved or distinguished from one another. Image quality can also be thought of as image sharpness, crispness, or contrast.

As discussed earlier, imagery is never perfect. It is limited by geometrical aberrations, diffraction, the effects of manufacturing and assembly errors, and other factors. The characterization of image quality by the methods described in the following sections will help you to assess just how your system performs with respect to its imagery.

It is important to realize that the image quality or resolution of the entire system is not totally dependent on the optics, but may include the sensor, electronics, display device, and/or other system components

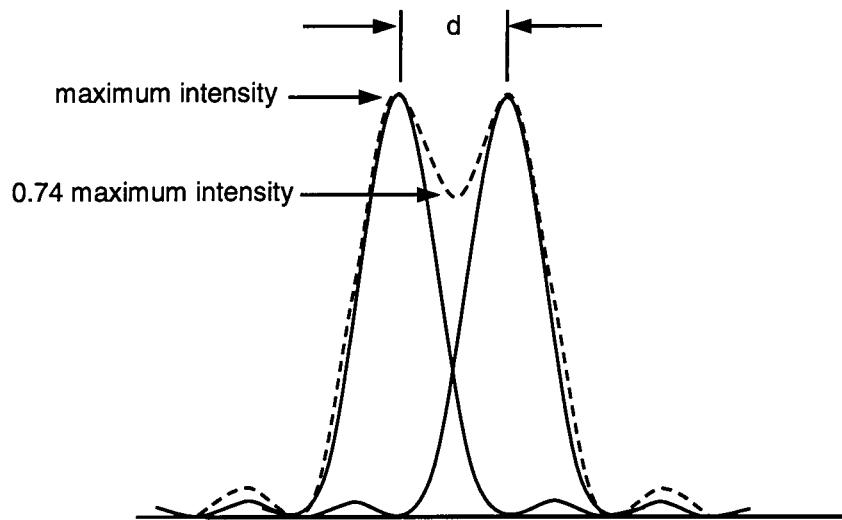
making up the system. For example, if the eye is the sensor, it can accommodate for both defocus and field curvature, whereas a flat sensor such as a CCD cannot. In this chapter, we will be discussing only the optics contribution to image quality.

What Is Resolution?

When we think about the image quality of an image-forming optical system such as a camera lens, the first parameter that often comes to mind is *resolution* or *resolving power*. Classically, the ability of an optical system to separate two closely spaced point sources at the nominal object distance is generally considered to be the resolution.

Consider a perfect optical system which has an entrance pupil diameter, D and focuses to an image with a given $f/ \#$. Two point sources closely spaced will be imaged through the optical system, each of them forming a diffraction pattern. If two perfect diffraction patterns as in Fig. 10.1 are separated by the radius of the Airy disk (the radius of the first dark ring in the diffraction pattern), then the intensity midway between the two peaks in the pattern drops to 0.74 of the maximum intensity, and the two point images are said to be *resolvable*. This is the Rayleigh criterion

Figure 10.1
Two Resolvable
Images of Closely
Spaced Point Sources



for resolution. This, of course, assumes that the ultimate media or sensor is not the limiting factor. The separation, d , of the two points in the image plane is

$$d = 1.22 \lambda f/\#$$

in units of wavelength. In object space, in radians, this becomes (λ and pupil diameter in same units)

$$\alpha = \frac{1.22 \lambda}{\text{entrance pupil diameter}}$$

or very frequently used as a rule of thumb, the resolution, in arc seconds, for the visible spectral range is

$$\alpha = \frac{136}{\text{entrance pupil diameter}}$$

where the entrance pupil diameter is given in millimeters. This is interesting and certainly of value in understanding the limiting resolution of the optical system with given first-order optical parameters, but it really does not help us to understand the performance of a specific optical system design. As will be shown in this chapter, there is far more to the characterization of optical performance than the theoretical resolution.

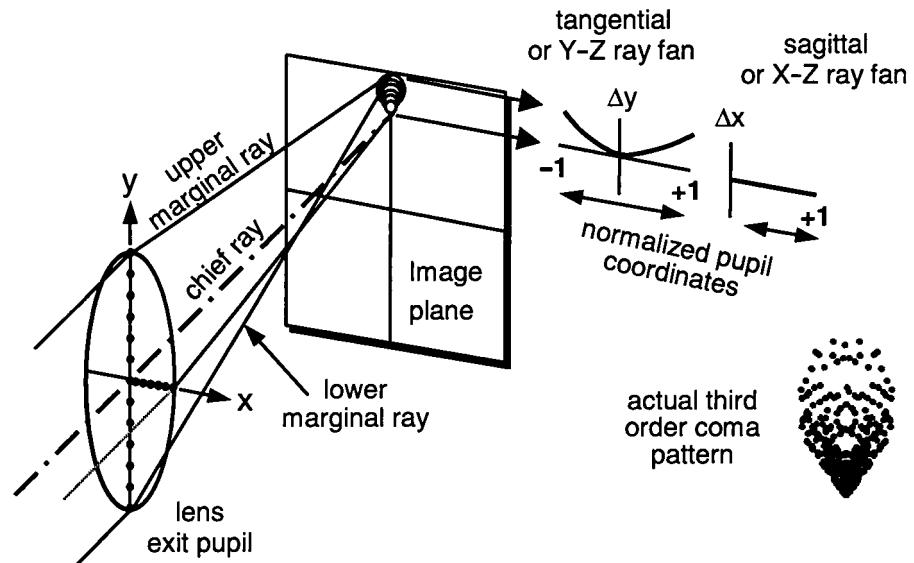
Ray Trace Curves

Most of the methods used in computing image quality, such as the modulation transfer function, spot diagrams, encircled energy, and the like, are functionally robust and represent different, yet similar representations of the net performance of the optical system as designed. However, there are two disadvantages with these metrics. First, they can sometimes take too much time to compute. This, however, is less and less of a problem as PCs have become faster and faster. Second, the real problem is that while these metrics do help to show the overall net resulting image quality, they do not provide a detailed indication to the designer of the specific aberrations present in the design over the field of view and over the spectral bandwidth. While some information can at times be derived, more often the user really cannot tell what aberrations are present and at what magnitudes. These data are important to the designer as an aid in correcting the residual aberrations.

The solution is to generate what are called *transverse ray aberration curves* or simply *ray trace curves*. With these graphical data, a reasonably experienced designer can immediately tell just how much spherical aberration, coma, astigmatism, field curvature, axial color, lateral color, and field curvature are present. In addition, in many cases the user can also tell what orders of these aberrations are present. Finally, with this knowledge, the designer can often make a reliable judgment as to what to do next regarding further optimization of the lens. In spite of some fabulous advances in performance simulation and modeling, transverse ray aberration curves are still invaluable to the serious designer.

We show in Fig. 10.2 the basic formation of the ray trace curve. This perspective figure (Fig. 10.2) shows a lens exit pupil with the lens imaging to an off-axis image position. First, consider tracing the *chief ray* to the image. The height on the image of the chief ray is our reference point, and is generally taken to be the image height. Now let us trace a ray through the top of the exit pupil. This ray, which is called the *upper marginal ray*, hits the image higher than the chief ray for the aberration shown, which is coma. Now let us trace a ray through the bottom of the exit pupil. This ray, which is called the *lower marginal ray*, also hits the image higher than the chief ray, and in fact for classical third-order coma it hits the image the same distance above the chief ray as the ray from the top of the pupil. In

Figure 10.2
Explanation of Ray
Trace Curves



other words, both of the rays from the top and the bottom of the exit pupil hit the image vertically displaced by the same amount.

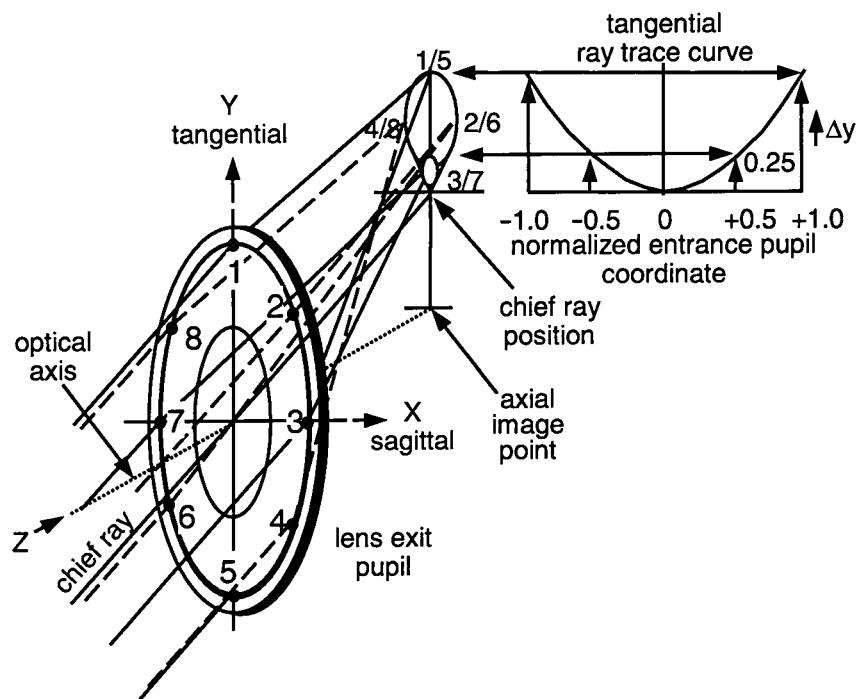
We will now proceed to establish a set of coordinate axes for our ray trace curves. In the first set of coordinates (on the left in Fig. 10.2), the abscissa is the normalized exit pupil radius in the y direction, and the ordinate is the distance above or below the chief ray on the image that our ray intersects the image plane (Δy). Thus, both the upper and lower marginal rays form the end points on the curve. We now proceed to trace rays through each of the black dots from $y = +1$ to $y = -1$, with the intersection points relative to the chief ray plotted on the curve. For third-order coma the result will be a quadratic or parabolic curve since third-order coma is quadratic with aperture.

Now we establish a second set of coordinate axes, as shown on the right in Fig. 10.2. Here we have the normalized X coordinate in the exit pupil in the abscissa, and the displacement of the ray in the x direction (Δx) as the ordinate. As it turns out, for third-order coma there is no x departure at all for these rays in the exit pupil. We will show why this is the case shortly. For now, you will see that for third-order coma a quadratic curve in the “tangential” ray fan and zero departure for the x rays in the “sagittal” ray fan are the results. If any lens designer who is “worth his or her salt” sees this form of ray aberration curves with a quadratic in the tangential ray fan and virtually zero in the sagittal curve, then he or she should conclude instantly that the lens has third-order coma.

Since these ray trace curves are so fundamentally important to the optical designer’s work, a more in-depth discussion is in order. As you will see, there are here, as with many other areas of optical design, subtleties that could easily be misleading if not fully understood. Consider our coma pattern where the ray trace curves suggest zero x departure of the rays hitting the image, which implies or suggests zero x width to the image blur. Yet we all know that coma does have width in the x direction. Just what is going on, and why are the data misleading?

Figure 10.3 will explain the situation. Here we trace rays around the periphery of the exit pupil from positions 1 through 8. From our prior discussion, we know that the chief ray is our reference, and that rays 1 and 5 from the top and bottom of the exit pupil both hit the image high, above the chief ray. If you follow the numbers in Fig. 10.3, you will see how *one* rotation around the exit pupil results in *two* rotations around an ellipse in the image, and since positions 1 and 5 are both high, then positions 3 and 7 which are 180° opposed will be at the bottom of the elliptical pattern. Neither of these rays will have any x departure at

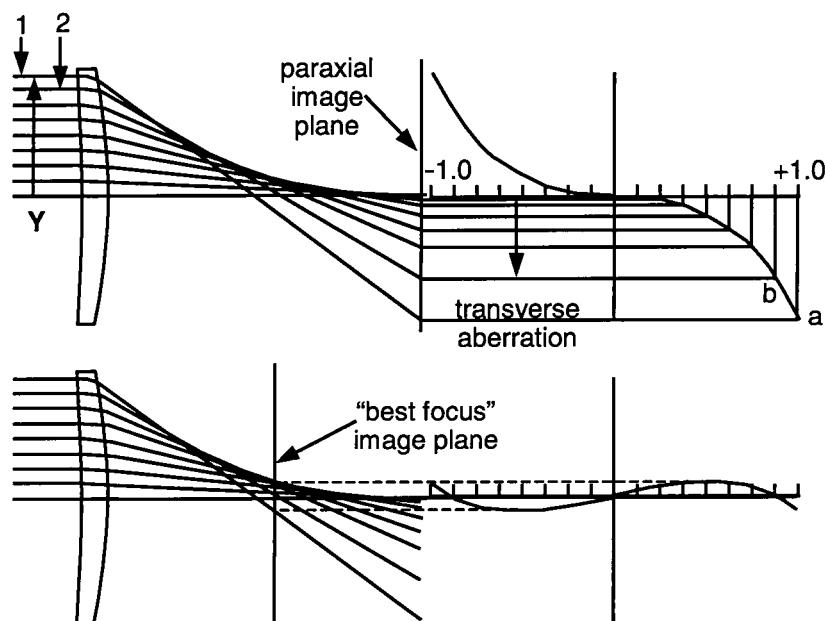
Figure 10.3
Formation of Comatic
Image Blur



all! Thus, the ray trace curve for rays traced in the x direction was a horizontal line in Fig. 10.2. So where is the x spreading of the coma pattern coming from? The answer is from rays at positions 2, 4, 6, and 8, which are called *skew rays*. Since the rays making up the ray trace curves contain only the y (or *tangential rays*) and the x (or *sagittal rays*), the designer sees no indication or evidence whatsoever of the x spreading of the imagery. This is a real subtlety, and it is a fine example why one should never be totally dependent on only one form of image evaluation or analysis. By looking only at the ray trace curves, one could easily conclude that such a system had virtually zero x spreading of the off-axis imagery, and this could make its performance ideal for some system applications. For the most part use of the ray trace curves are wonderfully helpful and revealing; however, do be aware of subtleties as pointed out earlier.

A further illustration of the ray trace curves, Fig. 10.4 shows how spherical aberration is formed and how the ray trace curves are derived. In the top of Fig. 10.4 the image is located at paraxial focus and it should be clear how each of the rays entering the lens from the left results in a

Figure 10.4
Formation of Ray
Trace Curves for
Spherical Aberration



corresponding intercept on the image plane and how this is plotted as the ray trace curve. Ray 1 strikes the image lowest and results in point *a* in the plot. Ray 2 is the next ray lower down entering the entrance pupil and it results in point *b* in the plot, and so on. Since third-order spherical aberration is cubic with aperture, the resulting curve is cubic. Note the symmetry above and below the optical axis. Now consider what happens if we relocate the image plane to the “best focus” position. Following the same logic in generating the ray trace curves, we see a much lower departure of the ray intercept points making up the curve. This is true and quite real, and it tells us that the image blur diameter when we refocus the image will be significantly reduced from that at paraxial focus. As an exercise, what will the ray trace curve be for a perfect image where the image plane is intentionally defocused toward the lens? The answer is a straight line sloped upward to the right. So let’s use this as an aid in further understanding ray trace curves. Since defocus yields a sloped but otherwise straight line, we can easily determine what any ray trace curves for any lens will look like as we go through focus by simply drawing or imagining a sloped straight line as a new coordinate axis. This is an invaluable tool as you can now immediately assess the relative improvement after refocusing a given lens. And since field curvature is a quadratic

change in focus with field of view, you can with a little practice assess immediately the benefits of curving your sensor if this is possible.

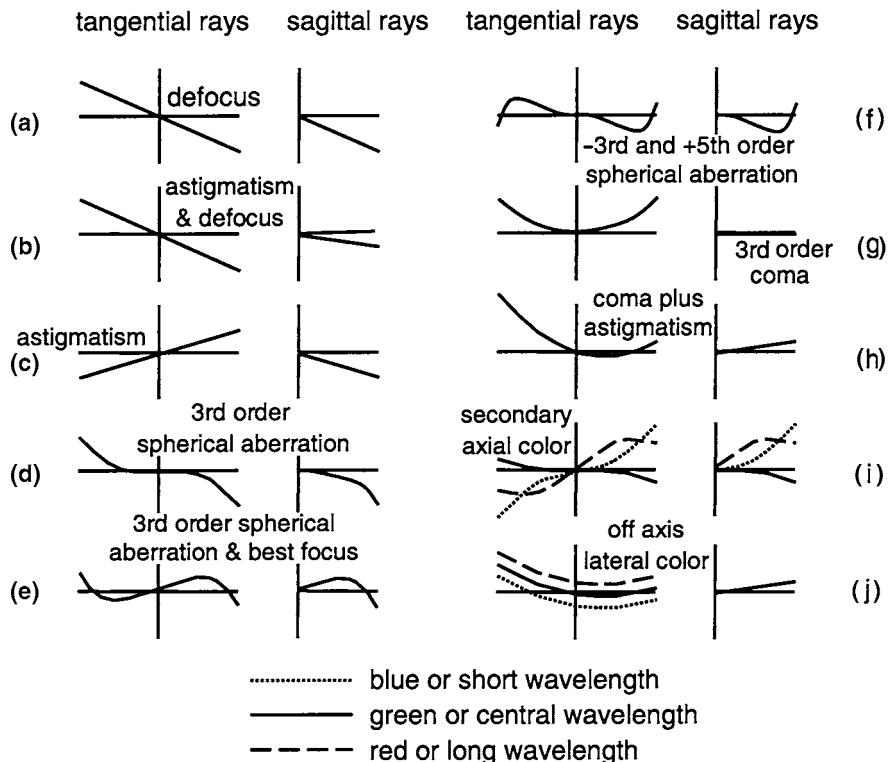
We show in Fig. 10.5 ray trace curves for various typical aberrations and combinations of aberrations:

Figure 10.5a is pure defocus. As noted earlier, defocus will produce equal sloped straight lines in the sagittal and tangential ray fans. Recall that the tangential ray fan is in the y - z direction and typically oriented parallel to the field-of-view direction. The sagittal ray fan is orthogonal to the tangential ray fan, and typically the sagittal fan is fully symmetrical which is why we sometimes show only one-half of the fan.

Figure 10.5b shows straight lines at different slopes. This is a combination of astigmatism (which is the difference between the slopes of the two curves) and defocus.

Figure 10.5c shows that if we best focus for the residual astigmatism off axis as we might do with a curved image surface, we find the result

Figure 10.5
Typical Transverse Ray
Aberration Curves



here where an equal and opposite ray fan slope results in the tangential and sagittal directions.

Figure 10.5d is negative or undercorrected third-order spherical aberration, which is, of course, a cubic with aperture.

The data in Fig. 10.5e are the same third-order spherical aberration as in Fig. 10.5d, only we have refocused the image to a more optimum focus position to minimize the residual blur diameter.

Figure 10.5f shows negative third-order spherical aberration, which is being balanced by positive fifth-order spherical aberration.

The data in Fig 10.5g are for pure third-order coma, which, as we know from before, is quadratic with aperture. We also know from before that the sagittal curve indicates zero image blurring in the sagittal direction. This may be misleading and is due to the nature of coma formation and the fact that the ray aberration curves show only the rays along two lines in the pupil plane.

The data in Fig. 10.5h are for a combination of third-order coma and astigmatism.

Figure 10.5i shows a combination of some negative third-order spherical aberration at the central wavelength as well as secondary axial color and spherochromatism. The secondary axial color is the focus difference between the central wavelength and the common red and blue focus, which together are focused beyond the central wavelength. The spherochromatism is the change in spherical aberration with wavelength.

Finally, Fig. 10.5j shows an off-axis ray trace curve with primary lateral color or color fringing along with a small amount of coma and astigmatism.

It should be quite apparent that the ray trace curves for each of the aberrations has its own distinctive form, and this is what makes them so useful. The aberrations, in effect, add algebraically, so it is easy to tell almost immediately what aberrations are present in a given design at the different field positions.

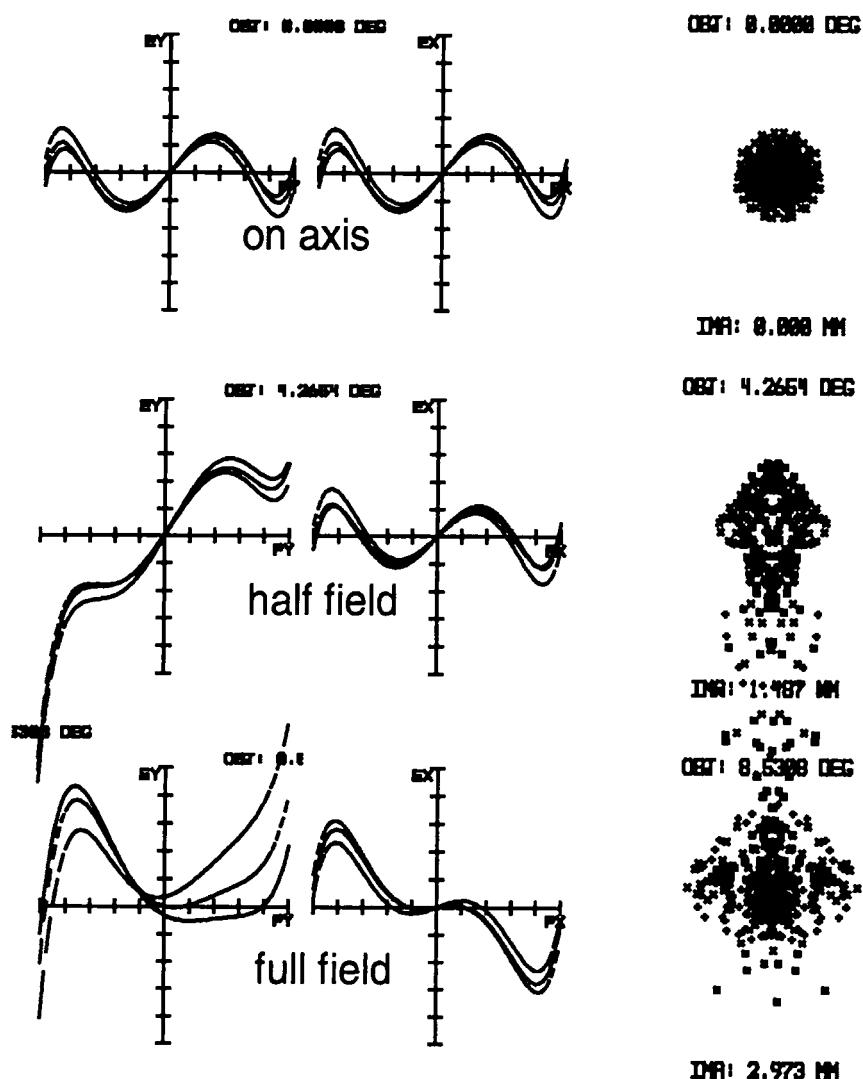
Spot Diagrams

Spot diagrams are the geometrical image blur formed by the lens when imaging a point object. This is a more functionally useful form of output; however, it is sometimes difficult to distinguish the specific aberrations present.

Figure 10.6 shows the spot diagrams for a Cooke triplet form of lens and shows both the transverse ray aberration curves as well as spot diagrams for the same field positions. Generally, the rms spot radius or diameter is output with the spot diagrams. The rms spot diameter is the diameter of a circle containing approximately 68% of the energy. This metric can be of great value, especially when working with pixelated sensors where one often wants the image of a point object to fall within a pixel.

Figure 10.6

Geometrical-Based
Spot Diagrams and
Transverse Ray Aberra-
tion Curves for
Cooke Triplet
Example



Note that in optical design software we often come across the terms “spot radius,” “spot diameter,” and “spot size.” While the designer is most interested in spot diameter, the software generally outputs spot radius. The use of the words “spot size” is fine for relative comparison (for example, “the spot size has increased by a factor of 2”); however, the term can cause undo confusion when tied to a specific value. For example, “the spot size is 50 μm ” does not really tell us whether this is the radius or diameter of the image blur, nor does it tell us whether this is for 100% of the energy or some other value such as the rms. Be careful in interpreting these forms of data from the software you are using.

Optical Path Difference

As we know from Chap. 4, if the peak-to-valley optical path difference (P-V OPD) is less than or equal to one-fourth of the wavelength of light, the image quality will be almost indistinguishable from perfect. It is known as diffraction limited. Just like transverse ray aberration curves, we can plot the optical path difference, and Fig. 10.7 shows a typical OPD plot for our sample Cooke triplet. In addition to the plotted data, we show also perspective views of the three-dimensional wavefront departure from perfect. Note how the curve and the perspective view correlate so well for the on-axis field position with the bump in the center as well as the turned up edge of the wavefront clearly evident in both data.

It is a little difficult without extensive experience to quickly determine the residual aberrations from an OPD plot, so you will generally want to compute the more standard ray trace curves for this purpose.

Encircled Energy

Encircled energy is energy percentage plotted as a function of image diameter. One good example of how we might use encircled energy is to specify an imaging optical system using a CCD sensor. Let us assume that the pixel pitch of our sensor is 7.5 μm . A good reliable and simple specification is that 80% of the energy from a point object shall fall within a diameter of 7.5 μm . Figure 10.8 shows an encircled energy plot for our sample Cooke triplet. Eighty percent of

Figure 10.7

Optical Path Difference for Cooke Triplet Example

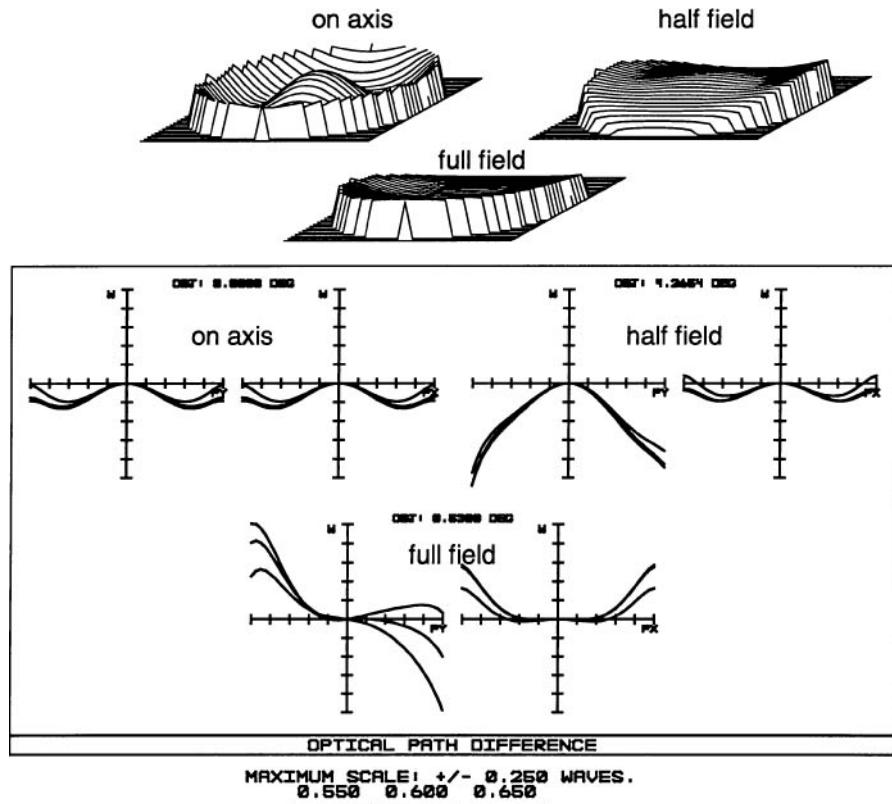
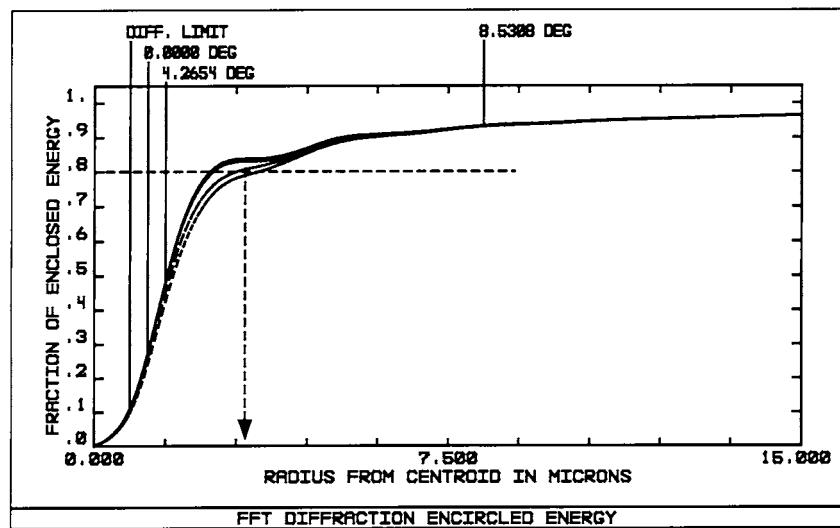


Figure 10.8

Encircled Energy for Cooke Triplet Example



the energy is contained within a diameter of approximately $6 \mu\text{m}$, which is a good match to the sensor. It also leaves some margin for manufacturing tolerances.

MTF

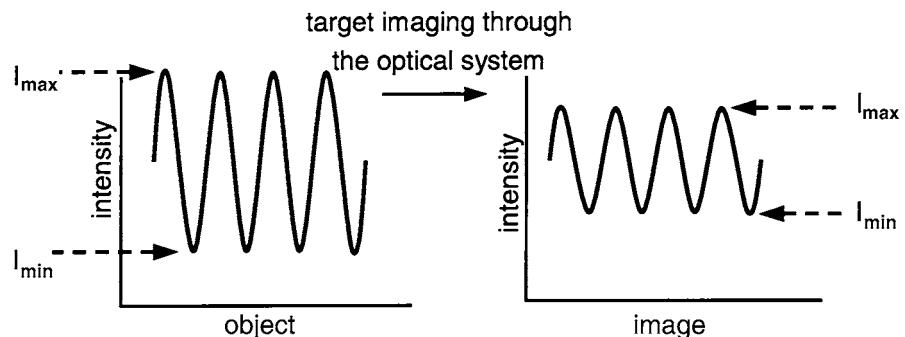
MTF is perhaps the most comprehensive of all optical system performance criteria, especially for image forming systems. Figure 10.9 is a representation of what is happening. We begin with a periodic object or target, which is varying sinusoidally in its intensity. This target is imaged by the lens under test, and we plot the resulting intensity pattern at the image. Due to aberrations, diffraction, assembly and alignment errors, and other factors, the imagery will be somewhat degraded and the brights will not be as bright and the darks will not be as dark as the original pattern.

Let us define some terms:

$$\text{Modulation} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad \text{MTF} = \frac{\text{modulation in image}}{\text{modulation in object}}$$

The modulation is simply the maximum intensity minus the minimum intensity divided by the maximum plus the minimum. The MTF is the ratio of the modulation in the image to the modulation in the object as a function of spatial frequency, which is generally in the form of line pairs per millimeter. Thus, the modulation transfer function represents *the transfer of modulation from the object to the image by the lens as a function of spatial frequency*.

Figure 10.9
Illustration of the
Meaning of the
Modulation Transfer
Function



There is another term, *contrast* (sometimes called *contrast ratio*), which is given by

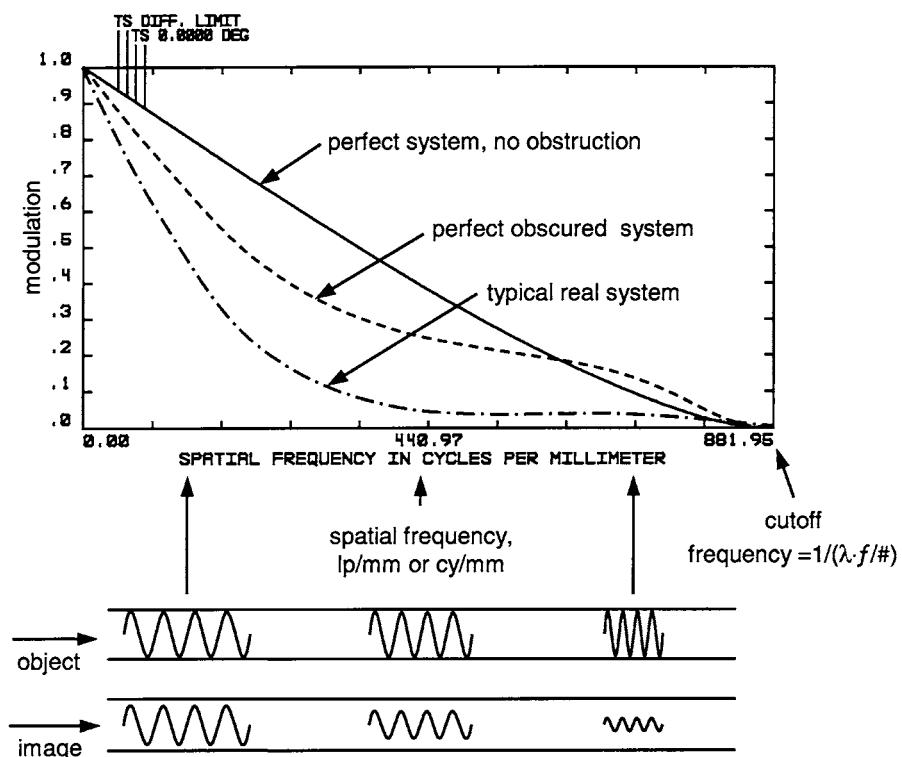
$$\text{Contrast} = \frac{I(\max)}{I(\min)}$$

Figure 10.10 shows several typical MTF curves. We show the MTF of a perfect optical system, a perfect system with a central obscuration (such as a Cassegrain telescope), and a typical real system. The MTF of the perfect obscured system has more diffraction due to its obstruction, and thus a lower MTF. The cutoff frequency, which is where the MTF goes to zero, is

$$v_{\text{cutoff}} = \frac{1}{\lambda(f/\#)}$$

The example shown is an $f/2$ lens in the visible ($0.55\text{-}\mu\text{m}$ wavelength), and the cutoff frequency is approximately 882 line pairs/mm. A good

Figure 10.10
Typical MTF Curves

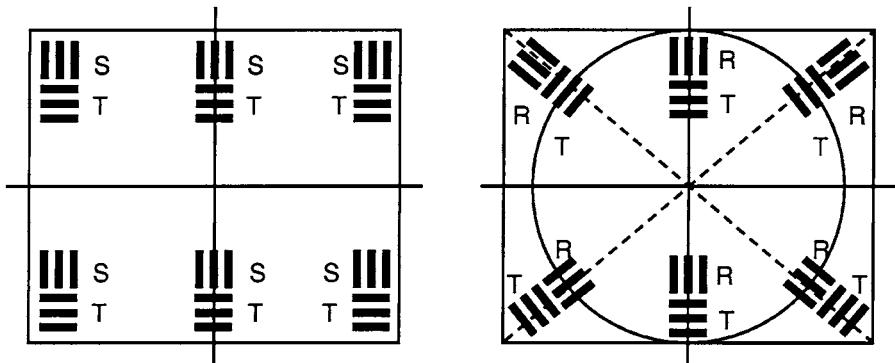


rule of thumb to remember is that an $f/2$ lens used at $0.5 \mu\text{m}$ has a cut-off frequency of 1000 line pairs/mm, and it is very easy to scale from here. For example, an $f/4$ lens has twice the Airy disk diameter and thus half the cutoff frequency of 500 line pairs/mm.

Recall that the MTF tells us how well the modulation in the object is transferred to the image by the lens. We show below the MTF curve in Fig. 10.10 a graphical representation of an object and of the resulting image at a low spatial frequency, a midspatial frequency, and at a high spatial frequency. Think of the low spatial frequency as being very large tree trunks with a bright sky between them, and the high spatial frequency as being tiny close tree branches with the same bright sky between them. The sky is the same for both and the darkness of the bark on the tree trunks and the branches is the same, hence the modulation of the objects is identical. However, the modulation of the image will be far lower at the higher spatial frequency of the branches since the MTF is lower.

The MTF is generally computed or measured for bars that are *radial* and for bars that are *tangential*. The radial bars are like the spokes on a bicycle, and the tangential bars are tangential to the edge of the bicycle tire as well as tangential to the edge of a circular field of view. The tangential and radial bars are orthogonal to each other. There is a subtlety of some significance with respect to the preceding nomenclature and how we handle it in the various software packages. Consider a lens designed to cover a rectangular format with a $3 \times 4 \times 5$ aspect ratio. We said earlier that the radial bars are parallel to spokes on a wheel and tangential bars are orthogonal to the radial bars. This is fine and true, and the definition is valid. However, given the pixelated nature of many of today's sensors such as CCD or CMOS detector arrays, or alternatively pixelated display devices as used in LCD and similar projection systems, it makes far more sense if the resolution bars and their associated nomenclature were consistent with the pixel rows and columns. What we mean is that it will be far more appropriate to refer to bars that are vertical and bars that are horizontal, rather than radial and tangential as shown in Fig. 10.10. Fortunately, most of the software packages use the terminology *sagittal* and *tangential*, and these refer to entrance and exit pupil coordinates. Tangential ray aberrations in the exit pupil are up and down in the plane of the paper in the y direction and sagittal ray aberrations are in and out of the plane of the paper in the x direction. We show in Fig. 10.11 the orientation of the bars using both nomenclatures. Regardless of which software package you use, it is imperative to understand what assumptions are being made with respect to target orientation.

Figure 10.11
Target Orientation Conventions



The bottom line here is as follows:

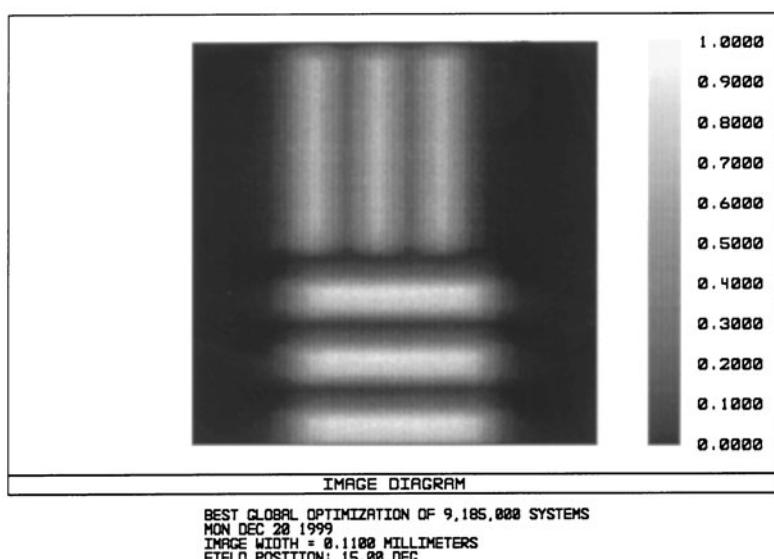
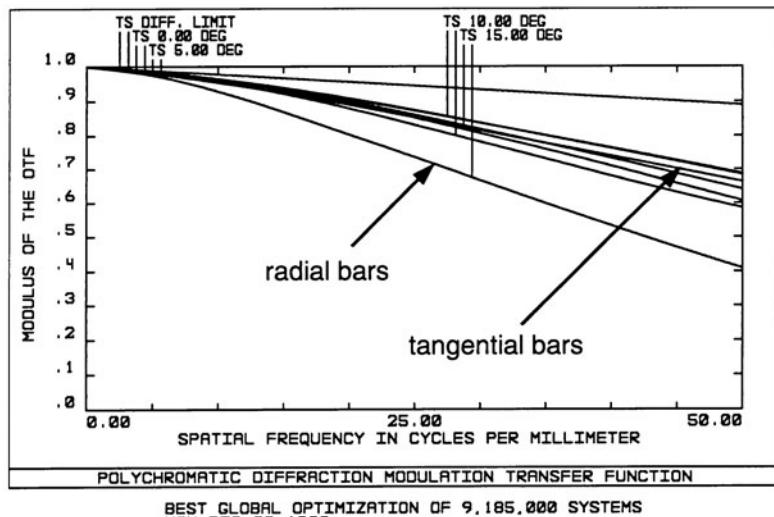
- Tangential aberrations are in the y direction and blur horizontal bars. Tangential bars are horizontal.
- Sagittal aberrations are in the x direction and blur vertical bars. Sagittal bars are vertical.
- Radial bars are bars parallel to spokes on a wheel.
- Tangential bars are tangent to the rim of a wheel.

A good example of a computed MTF and the resulting image quality is shown in Fig. 10.12. Here we plot the MTF of a double Gauss lens at four fields of view (on axis, 0.33 field, 0.67 field, and at the edge of the field). Note how the radial (vertical) bars at the edge of the field of view are more degraded than the tangential (horizontal) bars. The simulated 3-bar imagery is at 50 line pairs/mm, and the modulation at this spatial frequency is about 0.65 for the tangential bars and about 0.4 for the radial bars. The simulated imagery clearly shows this difference in the two target orientations. If you ever have a question with respect to this, we recommend that you read the manual for the software package you are using, phone the user support person, or, even better, run several examples.

Figure 10.13 shows the MTF of a perfect system as a function of obstruction ratio in a system such as a Cassegrain telescope. As the central obscuration increases, the amount of light that is diffracted away from the central maximum of the Airy disk increases and there is a corresponding reduction in MTF. As this is happening, the diameter of the central maximum actually decreases somewhat, resulting in a high-frequency

Figure 10.12

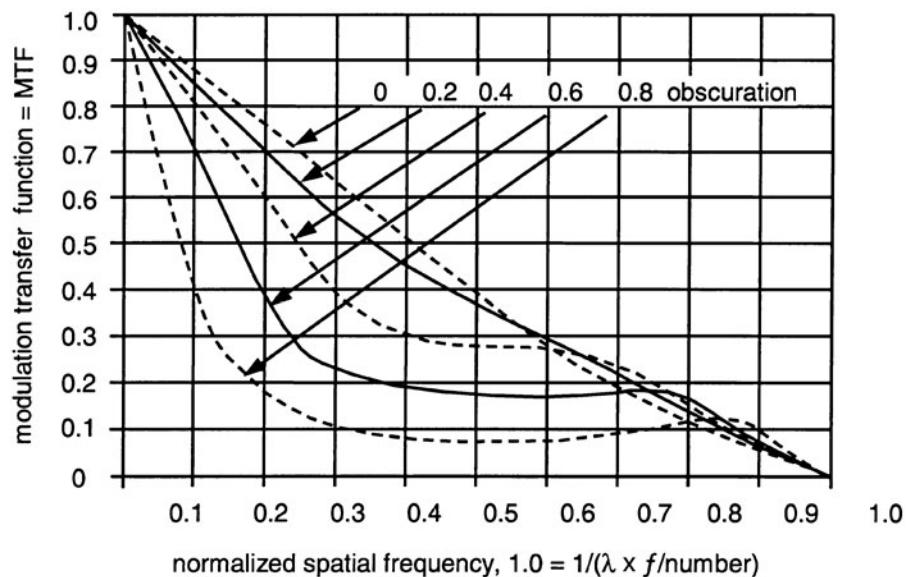
MTF and Simulated Bar Target Imagery of Double Gauss Lens



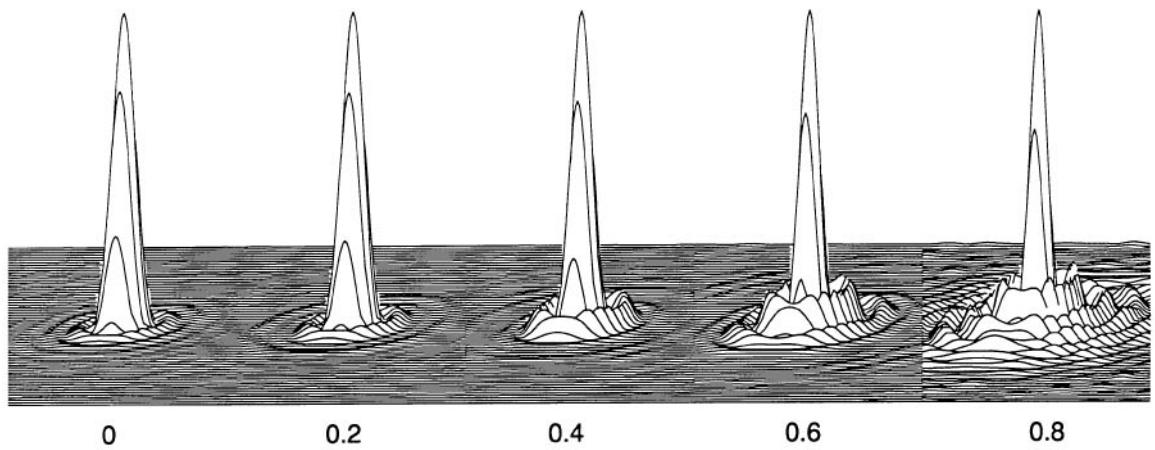
MTF that is actually slightly above the unobscured perfect system MTF. This can sometimes be used to advantage if you are trying to separate two close point objects or stars as in astrometric work. Figure 10.14 shows point-spread functions for aberration-free systems with no central obscuration, 0.2, 0.4, 0.6, and 0.8 diameter central obstructions, respectively.

Figure 10.13

MTF of a Perfect System As a Function of Central Obscuration

**Figure 10.14**

Point Spread Functions of an Aberration-Free System As a Function of Central Obscuration



It can be shown that the MTF of an image-forming optical system can actually be less than zero. As an otherwise well-performing system is degraded in performance due to defocus, aberrations, and/or manufacturing errors, the MTF degrades. As the performance continues to degrade, eventually the MTF may drop below zero. In extreme cases, the MTF will oscillate above and below zero. Any time the MTF is less than zero, that constitutes a phase reversal, which is where the dark bars become bright and the bright bars become dark. This can be seen visually by covering one eye and looking at Fig. 10.15. Hold the page at a distance closer than you can comfortably accommodate or focus. Do not try to focus on the pattern. This might range from 25 to 75 mm or more, depending on your accommodation. Then move the page closer or further from your eye very slowly, and you should be able to see very clearly the phase reversal. What is especially interesting here is what you are observing is the phase reversal of the imagery formed by your own cornea and eye as imaged onto your own retina... there is no other optics whatsoever! The phase reversal should be very clear and striking. If for some reason you cannot see it, Fig. 10.16 is a digital photograph of the effect. The camera lens was focused at infinity and located several inches from the figure to obtain the photo. Figure 10.17 shows the MTF

Figure 10.15

Radial Bar Pattern for Demonstration of Spurious Resolution

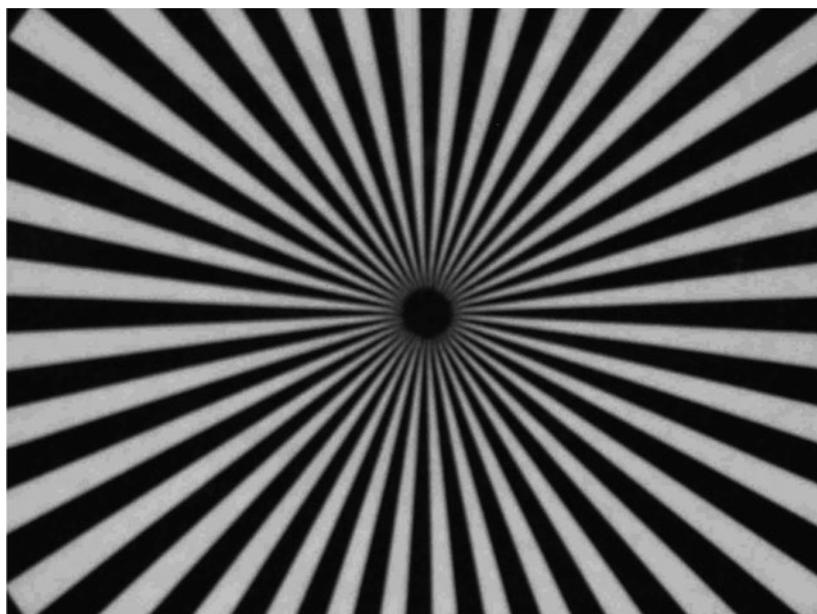
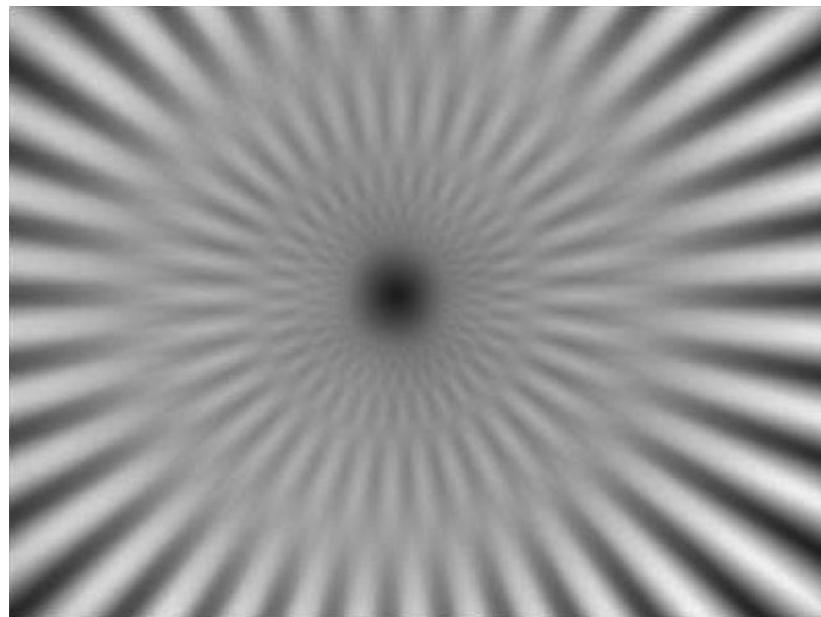
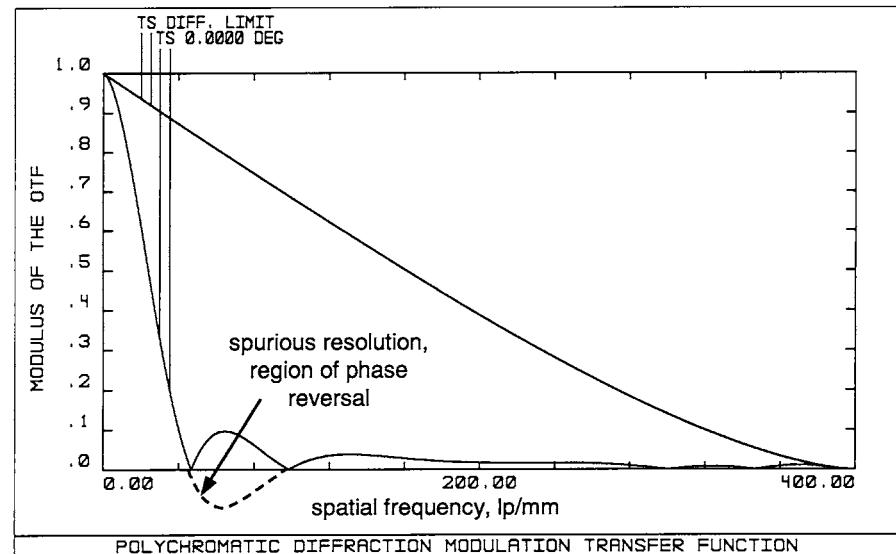


Figure 10.16
Spurious Resolution
Using Defocused
Digital Camera



of an otherwise perfect aberration-free $f/5$ system with approximately 1.5 waves of defocus. Note how at about 50 line pairs/mm the MTF is negative at about 0.1. In this spatial frequency range, there will be a phase reversal, as shown and discussed earlier.

Figure 10.17
Spurious Resolution
Due to Approximately
1.5 Waves of
Defocus

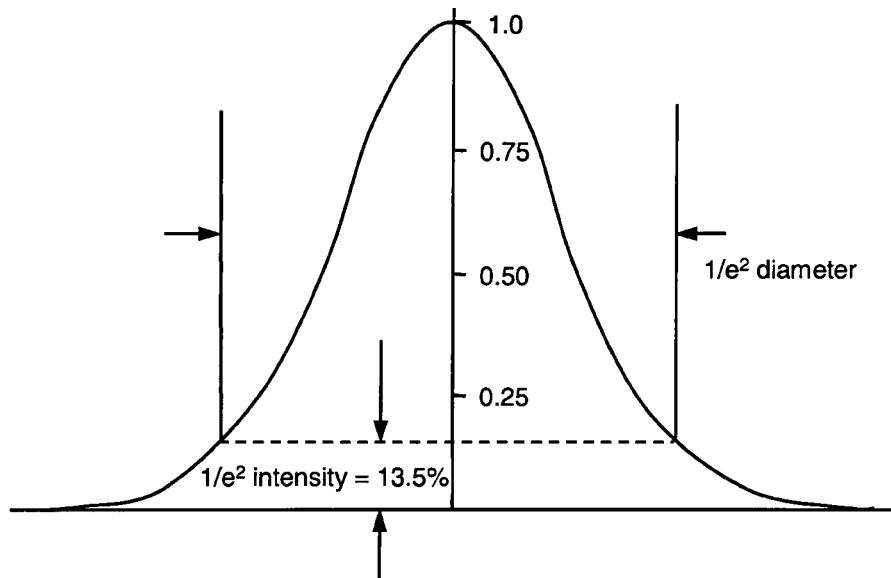


Gaussian Beam Imagery

Coherent light generated by lasers has properties different from light generated by other sources which we usually deal with in more conventional optical systems. If we look through a telescope at a distant object, the light intensity across the entrance pupil and aperture stop is uniform, and this is generally known as a “top-hat” intensity profile or distribution. A telescope objective, if it is free of aberrations, focuses a point object into an Airy disk pattern, with the diameter determined by the *f*/number or numerical aperture of the objective lens. In this case, a uniform top-hat distribution in pupil space transforms mathematically (by the optical system) to an Airy disk in image space. Laser beams emitted from rotationally symmetric resonators, such as HeNe or YAG lasers with a TEM₀₀ output, have an intensity distribution across the beam which is in the form of a gaussian intensity profile, as shown in Fig. 11.1. A gaussian intensity distribution in pupil space will mathematically transform to a gaussian in image space if the beam is not truncated by the aperture of the optical system, which is, of course, different from the uniform pupil transformation. Note that all of the material in this chapter assumes an aberration-free optical system. It is important to include the effects of lens aberrations in the final assessment of image quality and spot size.

The optical design of systems through which laser beams propagate is, therefore, very different from the design of conventional nonlaser systems used in either the visible or some other wavelength region. First, color correction of the laser-based optical system is much easier

Figure 11.1
Gaussian Intensity
Distribution



because the wavelength band of the laser light is extremely narrow. For HeNe lasers fully monochromatic light at $0.63282\text{ }\mu\text{m}$ can be used. Some lasers emit multiple spectral lines, or wavelengths can be changed. In these cases the optics must be designed to cover the functional spectral bandwidth. Laser diodes sometimes have a wavelength shift with temperature, which also must be taken into account by the optics.

Laser systems are often corrected for a small field of view because the laser beam enters the system parallel to the optical axis. Even if the real usable field of view is zero, the optical design must be optimized over a small, yet finite field of view in order to accommodate assembly and alignment tolerances. If one were to design such a system at identically zero field of view, it is possible that the performance may seriously degrade within 1 or 2 mrad off axis. However, there are systems, such as laser scanning systems, where some components of the system have to be designed for a large field of view. There are also systems where it is required to focus a laser beam to a very small spot, which requires the design of diffraction-limited optics with very large numerical apertures. In some cases, the laser beams have high-power densities, and they can cause damage to the optical components. Transparency of the material

from which the components are made can be degraded, and the material used for cementing of components and coating of the components can be damaged. The choice of optical materials is thus very important. In some cases, dust particles on the components in the system can also absorb enough energy to damage the surface of the component. Scattering from surface defects is a greater problem in laser systems than in visible and other incoherent systems.

The coherence length of a gaussian laser beam is large, and it appears as if the wavefront emerges from one point. If a gaussian beam is not truncated by the optical system, it emerges from the system as a gaussian beam. The narrow spectral line width of a laser beam and a well-defined wavefront permit very precise focusing and control of the beam.

Beam Waist and Beam Divergence

The beam emitted from a laser in TEM00, or fundamental mode, has a perfect plane wavefront at its beam waist position and a gaussian transverse irradiance profile that varies radially from the axis, which can be described by

$$I(r) = I_0 \exp\left(-2 \frac{r^2}{r_0^2}\right)$$

or by the beam diameter

$$I(d) = I_0 \exp\left(-2 \frac{d^2}{d_0^2}\right)$$

where I_0 is the axial irradiance of the beam, r and d are radius and diameter of a particular point in the beam, and r_0 and d_0 are the radius and diameter of the beam where irradiance is $(1/e^2)I_0$, or 13.5% of its maximum intensity value on axis. To define the propagation characteristics of a laser beam, the value, r_0 , is accepted as a definition of the radial extent of the beam. Finite apertures in the optical system or inside the laser itself, along with diffraction, cause the beam to diverge or converge. There is no such thing as a perfectly collimated beam without any spreading. Spreading of a laser beam is defined by diffraction theory. A laser beam converges to a point where the beam is smallest, called the *beam waist*, and it then diverges from this point

with a full angular beam divergence, θ , which is the same as the convergence angle. The beam divergence angle, θ , is the angle subtended by the $1/e^2$ diameter points in the far field, where the irradiance surface asymptotically approaches the full divergence angle, θ , shown in Fig. 11.2. At the location of the beam waist, the laser beam wavefront is flat, and it quickly acquires curvature on both sides of the waist. The beam waist diameter, d_0 , depends on the full divergence angle, θ , as

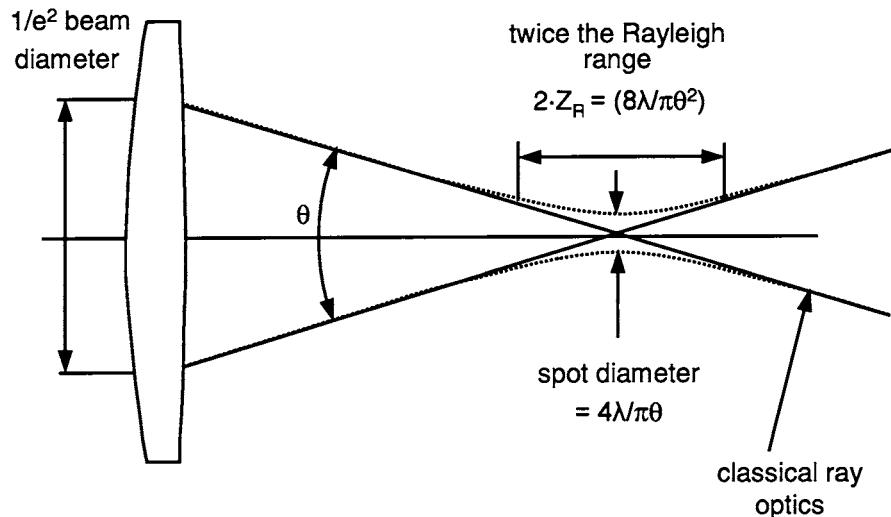
$$d_0 = \frac{4\lambda}{\pi\theta}$$

where λ is the wavelength of the laser beam and θ is in radians. It is important to know that the product of the divergence and the beam waist diameter is constant. The beam diameter grows with the distance, z , from the beam waist according to

$$d(z) = d_0 \sqrt{1 + \left(\frac{4\lambda z}{\pi d_0^2}\right)^2}$$

The radius of the wavefront at the beam waist location is infinite. As we move away from the beam waist, it then passes through a minimum at some finite distance, after which it rises again, approaching the value of

Figure 11.2
Divergence of a Laser Beam in the Far Field



z as z approaches infinity. The radius of the laser beam wavefront can be expressed as

$$R(z) = z \left[1 + \left(\frac{\pi d_0^2}{4\lambda z} \right)^2 \right]$$

The minimum value of the wavefront radius occurs at what is known as the *Rayleigh range*, where the beam diameter has the value $\sqrt{2}d_0$. At the extremes of the Rayleigh range, the image diameter is thus 41% larger than at the center of the beam waist. The Rayleigh range is sometimes used as a depth of focus number; however, do keep in mind that this reasonably large increase in spot diameter may not be acceptable. The Rayleigh range as measured from the beam waist is

$$z_R = \frac{d_0}{\theta} = \frac{4\lambda}{\pi\theta^2} = \frac{\pi d_0^2}{4\lambda}$$

The wavelength of the laser radiation, the beam waist diameter, the beam divergence angle, and the Rayleigh range are four parameters that completely describe a gaussian beam.

Collimation of Laser Beams

There is no perfectly collimated beam, since the beam divergence and the beam waist are defined by diffraction theory. It can be shown that the minimum beam spread between two points at a distance, z , happens when the starting beam waist is equal to

$$d_0 = 2 \left(\frac{\lambda z}{\pi} \right)^{1/2}$$

This relationship is equivalent to the one that gives the Rayleigh range, which tells us that the Rayleigh range is the distance inside which a gaussian beam has the minimum spreading. The diameter of the beam at the Rayleigh range is

$$d = \sqrt{2}d_0$$

If we use a beam expander at the output of a laser, we can increase the distance of the minimum spreading. If we additionally adjust the

beam expander so that the beam waist is located in the middle of the starting Rayleigh range, then the beam will spread to $\sqrt{2}d_0$ over a distance twice as long.

Propagation of Gaussian Beams and Focusing into a Small Spot

In 1983 S.A. Self derived a simple algorithm for tracing a gaussian beam through an optical system. The formula that he used had a very similar form to the paraxial lens formula that gives the relationship between the object position, focal length of a lens, and the image position. For a gaussian beam, Self calculates the Rayleigh range and the beam waist transformation by each lens in the system:

$$\frac{1}{s + [1/(s - f)]z_R^2} + \frac{1}{s'} = \frac{1}{f}$$

where s is the distance from the lens to the waist on the object side and s' is the distance from the lens to the new waist position. When the incident beam has its waist located in the front focal plane of a positive lens, the emerging beam has its waist at the rear focal plane.

Optical design programs generally use the gaussian beam propagation algorithm derived by A.E. Siegman. These programs usually calculate the radial beam size (semidiameter), the narrowest radial waist, surface coordinates relative to the beam waist, the semidivergence angle, and the Rayleigh range. This assumes the TEM00 fundamental mode, that is, a gaussian irradiance distribution. Lasers may be able to produce a number of other stable irradiance distributions, or modes, but they are not as compact as a gaussian beam, and they all have regions or holes in the beam, that is, where the irradiance drops to zero within the irradiance distribution. The mixed mode beams, defined by the beam quality factor, M , can also be traced in the current generation of optical design programs, in which case the M factor scales the embedded TEM00 gaussian mode.

In many applications, the goal is to focus the laser beam down to a very small spot. If the optical system focusing the beam is diffraction limited, the spot diameter at $1/e^2$ of the peak irradiance is defined by

$$d = \frac{4\lambda f}{\pi d_0}$$

The depth of focus is proportional to the square of the $f/\#$ of the focusing lens. If we define the allowable increase in the diameter of the focused spot, the depth of focus can be calculated using the formula for the spot diameter as the function of the distance from the waist location

$$d(z) = d_0 \sqrt{1 + \left(\frac{4\lambda z}{\pi d_0^2} \right)^2}$$

Truncation of a Gaussian Beam

Let us assume that we have a diffraction-limited lens with a given aperture diameter. The intensity profile of the focused spot is dependent on the intensity distribution of the radiation filling the aperture of the lens. For a uniformly illuminated aperture, the diffraction pattern is the classical Airy disk. It has a central bright spot and progressively weaker rings, with the first dark ring (intensity falls to zero) at a diameter, d :

$$d = 2.44 \lambda \cdot f/\#$$

When the illumination in the aperture is not uniform, the intensity profile in the focused spot does not have zero-intensity points, and the measure of the spot diameter is usually accepted as the diameter at which the intensity drops to $1/e^2$ of the peak intensity. When a gaussian beam falls onto the aperture of a lens, it may be truncated by the lens aperture. Let us define the truncation ratio as

$$T = \frac{d_0}{D}$$

where d_0 is a gaussian beam diameter measured at $1/e^2$ of the peak intensity and D is the aperture diameter of the lens. In the case of the lens aperture being two times the gaussian beam waist diameter ($T = 0.5$), the beam is truncated only below the intensity level of 0.03%. We can say that the effect of truncation is negligible, and the gaussian beam after transformation by the lens remains gaussian. On the other hand, when T , the truncation ratio, becomes a large number, and only the narrow central portion of the gaussian beam is transmitted through the lens, this case corresponds to the uniform illumination of the lens aperture, and the transmitted beam has the intensity distribution similar to the Airy disk. The beam waist diameter, d_0 , given in units of $\lambda \cdot f/\#$ for a few values of

TABLE 11.1

Beam Waist Diameter in Units of $\lambda \cdot f/\#$ for Several Values of the Truncation Ratio

Truncation Ratio T	d_0 at $1/e^2$ Intensity (in units of $f/\#$)	d_0 (Intensity Goes to Zero)	Truncation Intensity Level (%)
$\approx\infty$	1.64	2.44	100
2	1.69	—	60
1	1.83	—	13.5
0.5	2.51	—	0.03

truncation ratio is given in Table 11.1. Two cases of the intensity distribution for $T = \infty$ and $T = 1$ are shown in Fig. 11.3.

In Fig. 11.4, we show how the truncation ratio affects the performance of an otherwise perfect system with a gaussian intensity profile incident onto the entrance pupil of an imaging optical system. The abscissa is the ratio of the physical aperture to the $1/e^2$ beam diameter. The ordinate on the right is the normalized spot radius (normalized to $1/e^2$ spot radius).

Figure 11.3

Truncation of a Gaussian Beam and the Intensity Distribution at the Image Plane

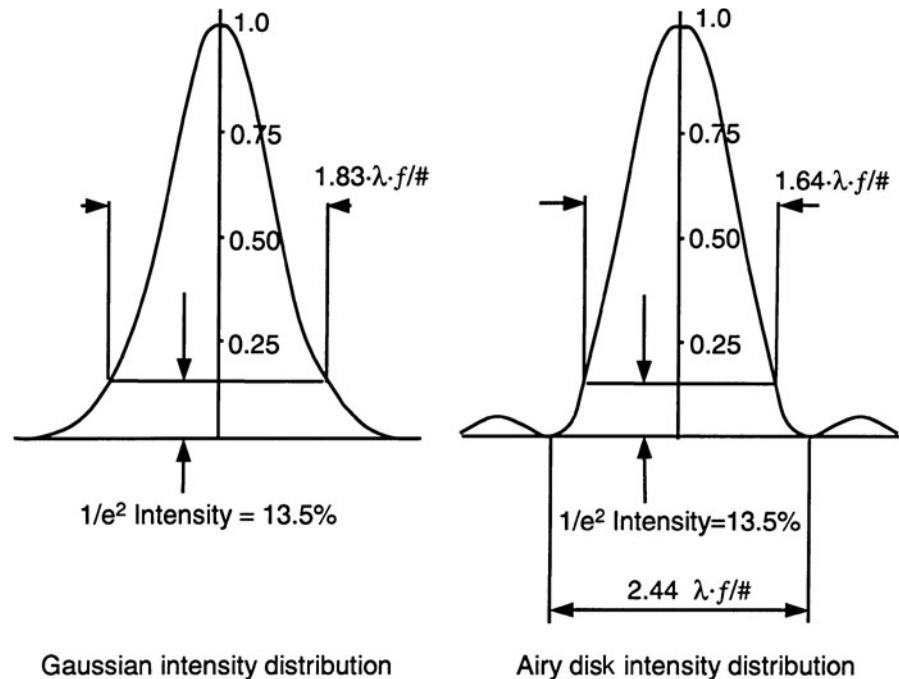
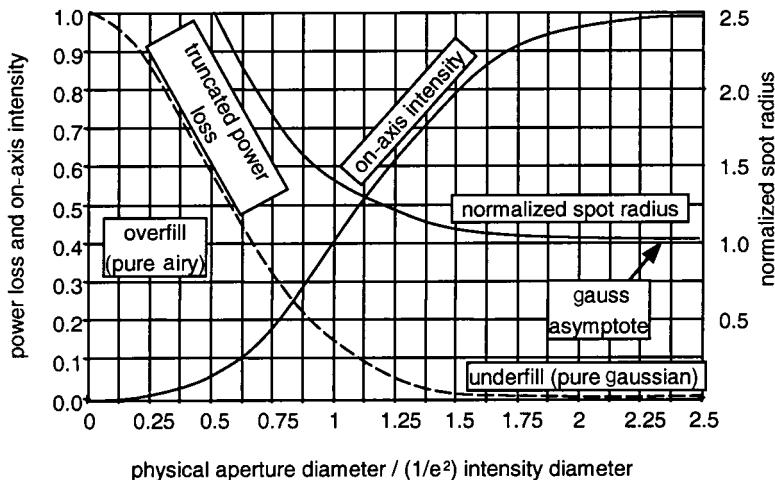


Figure 11.4

Truncated Gaussian Beam Imagery. Aperture Changes Relative to a Constant $1/e^2$ Diameter

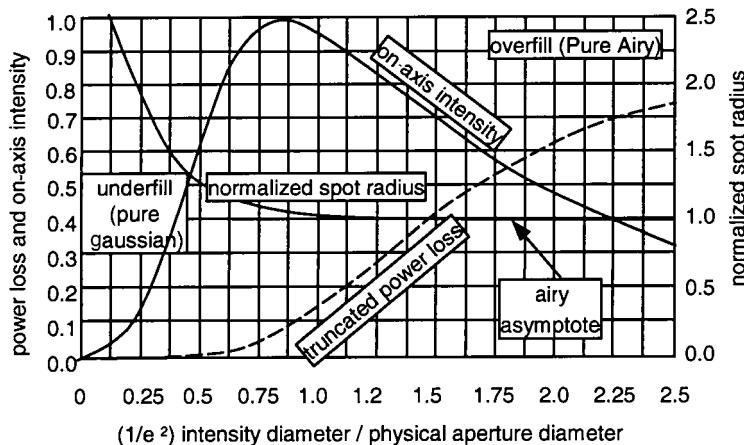


The easiest way to understand this somewhat complex data is to think of a constant $1/e^2$ beam diameter of, say, 25 mm. If the physical aperture diameter were 1 m, there would be virtually zero truncation of the beam, and a gaussian intensity distribution in the entrance pupil would transform into a perfect gaussian image profile at the image. The normalized spot radius asymptotically approaches unity as the physical aperture diameter continues to increase. The far right data point in the abscissa in Fig. 11.4 is for the physical aperture being 2.5 times the $1/e^2$ beam diameter. If, on the other hand, the physical aperture diameter were in the order of 3 mm for our 25-mm $1/e^2$ beam diameter, the ratio of the physical aperture diameter to the $1/e^2$ beam diameter would be about 0.125 and the intensity profile would be nearly a top-hat, which means that we will acquire diffraction and the imagery will be close to a classical airy disk pattern. It should be clear that if we truncate the beam at the $1/e^2$ beam diameter, the $1/e^2$ diameter of the image will be approximately 40% larger than an untruncated beam.

The ordinate on the left in Fig. 11.4 is for both the power loss and the on-axis intensity, with the abscissa having the same meaning as before. As the ratio of the physical aperture to the $1/e^2$ beam diameter increases, the on-axis intensity increases, and the loss in power due to truncation decreases.

In Fig. 11.5 we show a similar plot, only here we have the $1/e^2$ beam diameter changing relative to a constant physical aperture diameter. The interpretation and rationale is the same as for the prior data.

Figure 11.5
Truncated Gaussian Beam Imagery. $1/e^2$ Diameter Changes Relative to a Constant Aperture Diameter



There are several important messages from Figs. 11.4 and 11.5. First, a small beam truncation will affect the $1/e^2$ beam diameter of the resulting image. Truncating the beam to the $1/e^2$ beam diameter will increase the $1/e^2$ spot diameter by approximately 40%. In order to increase the spot diameter by less than 10%, we require a physical aperture about 50% larger than the $1/e^2$ beam diameter. The other data which are of significance are the intensity and power loss associated with a given truncation factor. If our physical aperture were 50% larger than the $1/e^2$ beam diameter, our on-axis intensity would be 80% of that from an untruncated beam.

Application of Gaussian Beam Optics in Laser Systems

Gas lasers have their place in many applications in the large consumer market. HeNe lasers, which emit red light at $\lambda = 632.82$ nm, are used in bar code readers, the printing industry, machine vision, etc. They are used wherever the packaging constraints are not too tight and the use of visible light is convenient. The beam out of the HeNe laser is TEM00 rotationally symmetric gaussian beam, which can be easily transformed with the properly designed optics to the required spot size.

Solid-state lasers can generate very powerful beams. They can emit TEM00 gaussian beams. There are a significant number of types of solid-state lasers that emit in the near IR spectral region. A YAG laser, which emits light at the wavelength of 1.064 μm , can be frequency doubled to produce green light at 532 nm. It can also be frequency tripled or quadrupled. YAG lasers are used in industrial applications where high power is needed, such as writing on metal, welding, cutting, hole drilling, etc. One disadvantage of solid-state lasers is that they are generally expensive. One field that makes an extensive use of lasers is medicine, especially in the area of diagnostics, cancer treatment, and eye surgery. The laser wavelengths used range from the UV below 200 nm to 10.6 μm in the far infrared. CO₂ gas lasers emit radiation at 10.6- μm wavelength, and they are widely used both in medical and industrial applications.

UV lasers are used in lithography to achieve the submicrometer imagery needed to make integrated circuit chips. The minimum spot size is determined by the wavelength and the numerical aperture of the focusing optics. The shorter the wavelength, the smaller the spot to which the laser beam can be focused. This pushes the microprocessor industry to use increasing shorter wavelengths. Currently, a common wavelength is 193 nm.

Laser diodes are undergoing an extremely rapid development. Advantages of laser diodes include their low cost and small physical size. They are also available in a wide range of wavelengths. Laser diode devices provide continuous output power or may be analog or digitally modulated. Pulsed laser diodes typically operate with pulses shorter than 100 ns. Laser diodes can be packaged in the form of a linear diode array or a two-dimensional array. Laser diodes may be temperature tuned by approximately 0.3 nm/ $^{\circ}\text{C}$. Applications of laser diodes range from telecommunications, data communications sensing, thermal printing, laser-based therapeutic medical systems, and, satellite telecommunications to diode pumping of solid-state lasers.

The primary optics-related disadvantage of laser diodes is that they emit nonrotationally symmetric beams, which are more difficult to collimate or focus into a small spot. Most beams from laser diodes have a gaussian intensity profile along one axis, often called the *fast axis*, and a nongaussian intensity profile along the *slow axis* perpendicular to it. The beam diverges much faster in the fast axis than in the slow axis. Achieving collimation of the beams from laser diodes is not a trivial task because of the lack of the rotational symmetry in the beam. It requires

anamorphic optics, in most cases treating the fast and the slow axes separately. A good way to collimate a beam from a laser diode is to use two crossed glass or gradient index fibers, perpendicular to the optical axis of the laser beam shown in Fig. 11.6, each fiber having a different focal length. This is an elegant solution because of the small packaging. Two perpendicular fibers have small focal lengths. The fiber that collimates the fast axis has a focal length of a few hundred micrometers, and can be a gradient index fiber with a radial gradient index distribution, or a planohyperbolic homogeneous fiber. The fiber that collimates the slow axis has a longer focal length. In this way, the output of most laser diodes can be transformed into near rotationally symmetric beams. The beam can then be focused with a rotationally symmetric lens into a single-mode or a multimode fiber core that transmits the laser signal along the fiber axis.

A very convenient way of transmitting laser beams is through fibers. The laser beam from a laser diode can be coupled into a fiber with a coupling efficiency as high as 90% and transformed to a clean beam, easier to collimate at the output of the fiber. In telecommunications, the laser light-carrying signals are transmitted through hundreds of kilometers of fiber.

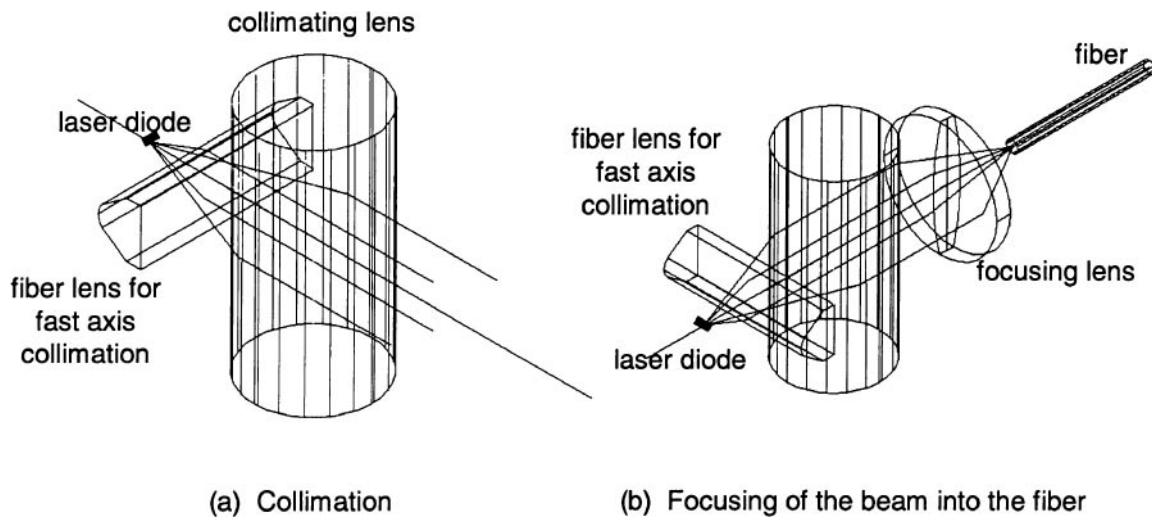


Figure 11.6

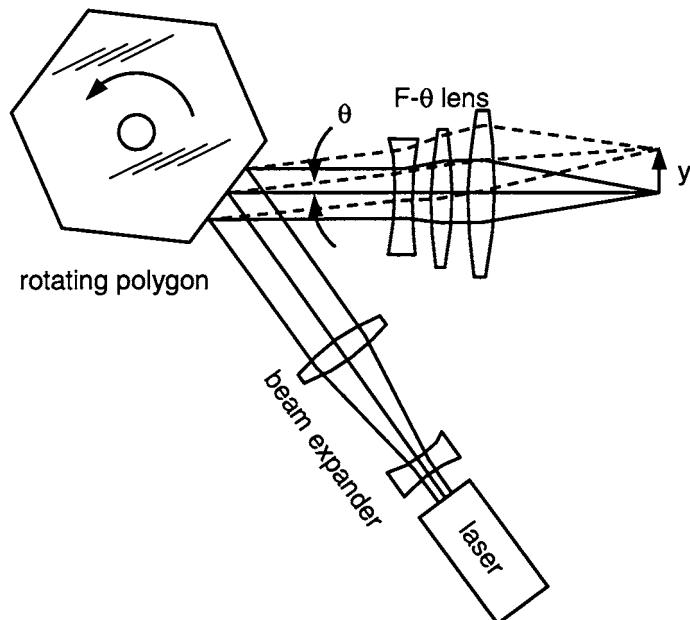
Collimation and Focusing of the Beam from the Laser Diode

F-θ Lenses in Laser Scanners

Laser scanners represent one of the more common applications of gaussian beam optics, and a very special lens form is required in these laser scanning systems. Figure 11.7 shows a very generic laser scanning system. An HeNe or similar laser beam is first expanded by a beam expander, and the expanded laser beam is directed toward a multifaceted polygon scan mirror. As the mirror rotates at a constant rotational speed, angle θ changes linearly.

Most lenses follow the relationship that the image height $Y = f \tan \theta$. This is literally true for lenses with zero distortion. If such a lens were used in our laser scanner, then the velocity of the scanned laser spot would increase in proportion to the tangent of the scan angle. Since most laser scanners require a linear spot velocity, conventional lenses are not viable. Lenses which follow the relationship $Y = f \theta$, will produce a linear scan velocity, and these lenses are called *F-theta*, or *F-θ*, lenses. In effect, they are lenses in which negative or barrel distortion is intentionally introduced in order to counteract the increased image

Figure 11.7
Laser Scanner with
F-θ lens



velocity in conventional low-distortion optics. Fortunately, these forms of remote-aperture stop lenses (the stop is on the polygon facet) tend to inherently have negative distortion. It is critical, however, to assure uniform spot size through scan, and this is sometimes challenging. In order to loosen the tolerance on pyramidal error of the polygon, the collimated beam from the laser is often focused with a cylindrical lens to a line on the polygon facet. This requires an anamorphic $F\text{-}\theta$ lens to focus the beam to a spot. This form of optical system is used in laser printers.

Basics of Thermal Infrared Imaging in the 3- to 5- and 8- to 12- μm Spectral Bands (Plus UV Optics)

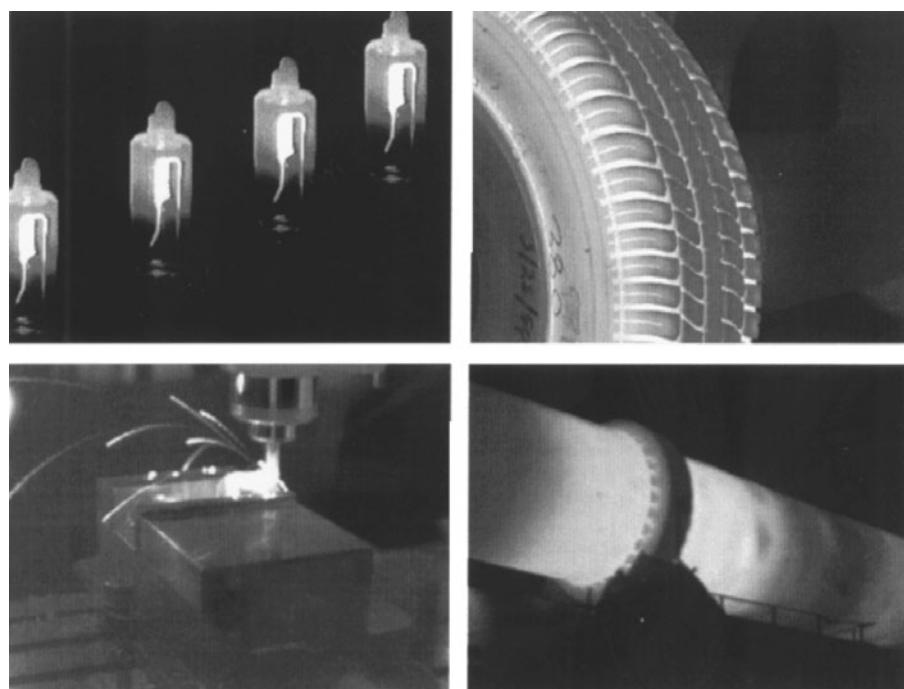
The Basics of Thermal Infrared Imaging

Thermal infrared imaging is generally considered to be in the medium or midwave IR (MWIR) that extends from 3 to 5 μm and in the long-wave IR (LWIR) that extends from 8 to 14 μm . In these wavelength bands, we are looking at thermal or heat sources rather than visible light. There are many different applications for thermal infrared imaging such as nondestructive testing whereby an IR camera can image a machine, such as a CNC lathe, to look for overheating of the bearings, or we can image houses looking for heat losses in the winter. In the medical arena,

doctors can look for various abnormalities indicated by localized skin temperature variations. In nuclear power plants IR cameras are invaluable to quickly search for thermal leaks in the cooling system. Boarder control and security are other areas where IR imaging has become crucial. Figure 12.1 shows several industrial and commercial examples of thermal infrared imaging. The applications of thermal infrared imaging are continuing to develop at a rapid pace.

The human eye is sensitive to the spectral band from approximately 0.4 to 0.7 μm , and thus the eye cannot see this longer-wavelength thermal energy. It takes special detectors or sensors to record the energy, and, needless to say, the imaging optics must efficiently transmit these wavelengths. While there are many applications for the near infrared (NIR), which includes the regions from 0.85 to 1.6 μm , for telecommunications as well as 1.06 μm , which is the Nd: YAG wavelength often used in the applications where the higher power is needed, for the most part ordinary optical materials can be used. The near IR will thus not be considered in this chapter, but rather the MWIR and LWIR where special optical materials and other design considerations are mandatory.

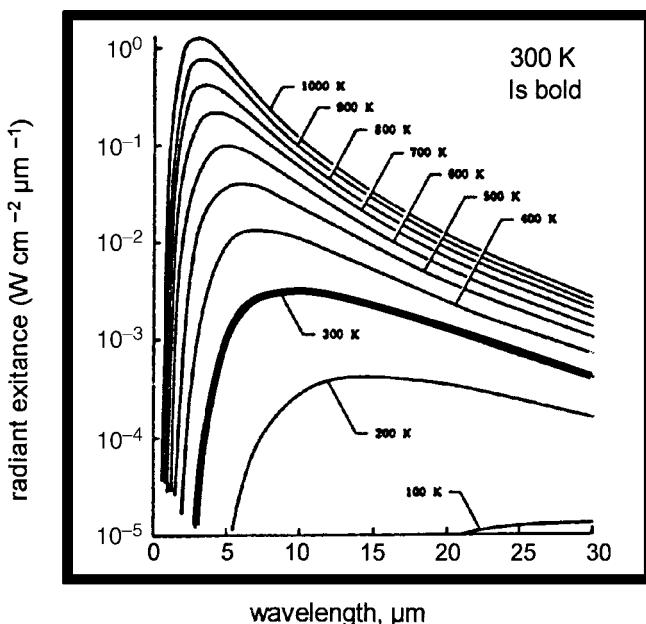
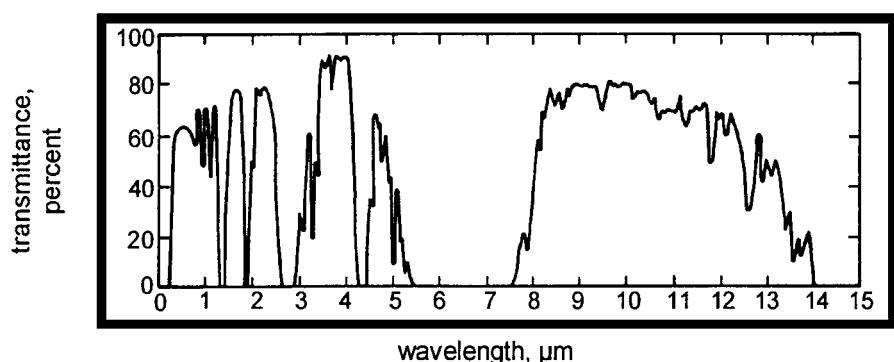
Figure 12.1
Examples of Industrial Applications of Thermal Infrared Imaging
(Courtesy of FLIR Systems, Boston)



We see in Figure 12.2 the spectral transmittance of a 1.8-km air path in the 3- to 5- μm MWIR and 8- to 14- μm LWIR spectral bands. Water and CO_2 absorption bands limit the use of wavelengths to these two bands within the atmosphere. We also see the radiant exitance for black bodies ranging in temperature from 100 to 1000 K. For reference, ambient temperature is about 300 K, and you can see that the peak radiant exitance at this temperature occurs at about 10 μm . For this reason, LWIR systems tend to have the highest sensitivity. However, LWIR detectors

Figure 12.2

Top: Transmittance of 1.8-km Horizontal Air Path at Sea Level;
Bottom: Radiant Exitance of Blackbody As Function of Temperature



are more expensive and difficult to produce than their MWIR counterparts. In addition, with today's image-processing hardware and algorithms excellent MWIR imagery can be obtained.

In a simplified form shown by Riedl, the signal-to-noise ratio for an IR system is

$$S/N = (W_T \epsilon_T - W_B \epsilon_B) (\tau) \left(\frac{D^*}{\sqrt{\Delta f}} \right) \left[\frac{\tau d'}{4(f/\#)^2} \right]$$

where ϵ = emissivity

W = radiant exitance (W/cm^2)

W_T = target exitance (W/cm^2)

ϵ_T = target emissivity

W_B = background exitance (W/cm^2)

ϵ_B = background emissivity

D^* = specific detector detectivity ($\text{cm} \cdot \text{Hz}^{1/2} \cdot \text{W}^{-1}$)

Δf = noise equivalent bandwidth (Hz)

τ = optical transmission

d' = detector size (cm) assuming square detector

$f/\#$ = $f/\#$ of optics

The first factor in the previous equation relates to the object we are imaging. This factor gives us the exitance difference between the primary object of interest or target and the area surrounding the object of interest or background. The second factor is the transmission of the atmosphere or other medium in which the system is immersed and the transmission through the optical elements. The third factor relates to the focal plane array and is the detectivity divided by the noise equivalent bandwidth. The fourth and final factor has the sensor size (the width of a pixel) and the optical transmittance in the numerator and the $(f/\#)^2$ in the denominator. This is where the optics becomes critical since the signal to noise is inversely proportional to the square of the $f/\#$. This drives many IR systems to extremely low $f/\#$ s in order to reach the desired signal-to-noise ratio. In addition, with some of the new uncooled microbolometers, the $f/\#$ often needs to be $f/0.8$ or even lower. The reader is referred to the reference for further information on this important relationship.

With respect to the optics, most glasses simply do not transmit above about $2.5 \mu\text{m}$. Certain special glasses do transmit up to $4.5 \mu\text{m}$, and fused

silica transmits up to about 4 μm . Infrared-transmitting materials are therefore essential, and as we will see there is only a limited selection available. These IR-transmitting materials are generally expensive and have other problems.

The Dewar, Cold Stop, and Cold Shield

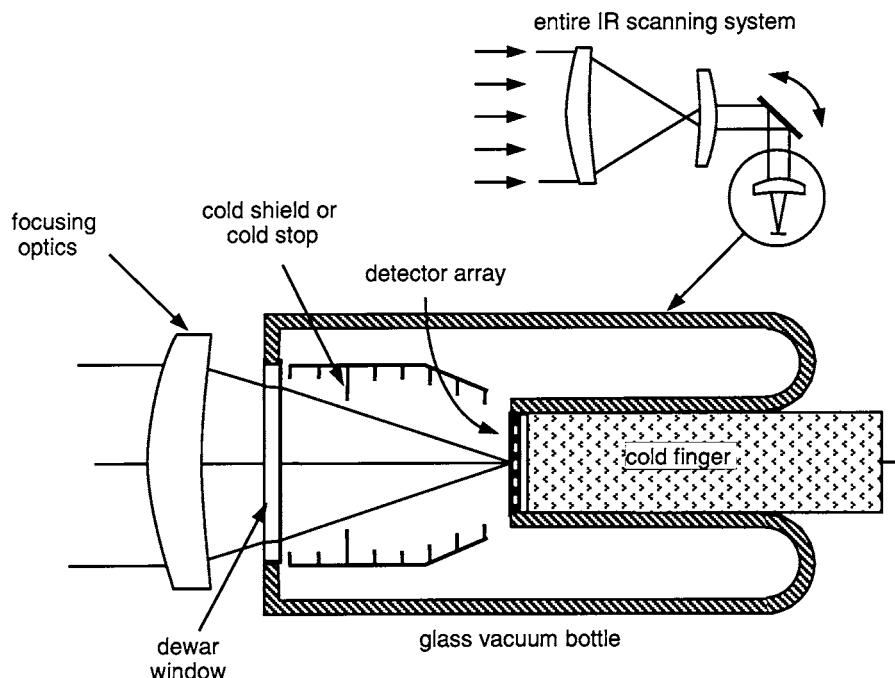
Since we are looking at thermal or heat sources with a thermal-imaging system, for maximum system sensitivity most thermal imaging systems use cryogenically cooled detectors which operate at the liquid nitrogen temperature of 77 K or even lower. If these detectors, or *focal plane arrays* (FPAs), are allowed to “see” any thermal energy other than the energy contained within the scene being viewed, then the sensitivity is reduced. In addition, if the magnitude of this nonscene energy changes or modulates over the field of view, then we often see cosmetically undesirable image anomalies. In order to achieve maximum sensitivity and avoid image anomalies, the IR FPA is cryogenically cooled and mounted into a thermally insulated “bottle,” or *dewar*, assembly.

Figure 12.3 shows a typical generic detector/dewar assembly intended for an IR-imaging application. Before we show how the dewar works, we need to see just how it interfaces with the rest of the optical system. The smaller figure on the upper right of Fig. 12.3 shows an entire scanning-imaging IR system. Light (actually infrared radiation) enters from the left into the larger lens generally called the *collecting optics*. After forming an intermediate image, the light is collimated by the second smaller lens. A further purpose of the second lens is to form an image of the larger collecting optics element, which is the system aperture stop, onto the scan mirror. After the light reflects from the scan mirror, it enters the region within the circle in Fig. 12.3. This area is shown enlarged and in more detail in the larger figure.

Collimated light from the scan mirror first enters the *focusing lens*, which is generally outside of the cryogenically cooled dewar and which focuses the light onto the FPA after passing through the dewar window. Note that for the example shown in Fig. 12.3 the detector array is a linear array extending in and out of the figure. The following attributes make up the dewar assembly:

Figure 12.3

Top: Entire IR Imaging System; Bottom: Typical Generic Detector Dewar Assembly



- The dewar is an evacuated bottle very similar to a classical Thermos bottle.
- The entrance window must, of course, transmit infrared radiation.
- A *cold finger* butts up against the aft end of the FPA to keep it cryogenically cold. The cold finger itself is a high-specific-heat metal rod of iron or steel, which is wrapped with a coil through which liquid nitrogen is pumped (or other similar operation). This, in turn, cools the aft end of the FPA.
- A baffle, called a *cold shield* or *cold stop*, is located as shown inside the dewar. We will define these later.

Now—an *extremely important point*—to evaluate what is happening with respect to the imaging light and potential stray light which may lead to undesirable image anomalies, you need to figuratively *put your eye at the detector and look out the front and ask yourself "what do you see?"* Let's do just that and put our eye figuratively at the center of the FPA where the light focuses.

- Within the solid angle forming the imaging cone of light, we “see” *scene energy*. We are, in effect, looking into the exit pupil of the optics, and we see a solid angle of radiation coming from the nominal scene. We do not see the image, but rather the exit pupil. This is much like looking toward a round porthole window in a ship from some distance away.
- At angles just outside the imaging cone, yet not quite hitting the cold shield baffle, we have radiation, which is not scene energy, nor is it cryogenically cold. This sliver of solid angle, which is a circular annulus, represents energy from the interior of the system, which literally reaches the FPA.
- Outside this sliver of interior system energy, we see the cold shield or baffle within the dewar. Since this component is cryogenically cold and finished with a nonreflective coating, it emits little or no radiation.

If, in the previous example, the cold shield had been the same diameter as the light cone, then the detector would only have been able to see scene energy. In effect, we have in this example made the cold shield slightly larger in diameter than it needed to be, and we will show why later. For now, we simply conclude that there is a sliver of solid angle, which is outside the scene energy and inside the cold shield, which can record energy from the system interior.

Cold Stop Efficiency

An IR system is said to be *100% cold stop efficient* if the detector can see or record energy only from the scene. What this really means is that with 100% cold stop efficiency, the detector records energy from both the cone of light representing scene energy and from the cryogenically cold thermal baffle, known as a cold stop (remember that being cryogenically cold, there is virtually no energy emitted from the cold stop itself). We have used the example in the previous section of putting your eye figuratively at the detector and looking out toward the front of the system and asking yourself “what do you see?” If, for every pixel on your FPA, you can convince yourself that your eye sees only the solid angle representing the imaging light (scene energy) and also portions of the thermal baffles representing the cold stop, then the system is indeed 100% cold stop efficient.

Note in Fig. 12.3 we have shown a series of small stray-light baffles within the cold stop or cold shield area. Without these stray-light suppression features, there may be stray radiation paths which will cause unwanted radiation to reach the FPA.

In Fig. 12.4 we show on the left a system which is not 100% cold stop efficient and one which is 100% cold stop efficient on the right. Casually looking at these figures shows little difference. The lower sets of figures are enlargements of the areas within the dashed circles of the upper figures. Note on the left figure how the aperture stop is both on the front element of the system as well as on the rear surface of the two-element reimaging lens group. Furthermore, if we place our eye at the lower end of the FPA as shown and look toward the scene, we see the solid angle representing the scene as well as a solid angle above this region and below the cold shield which is not coming from the scene but rather from some portion of the system interior. This nonscene energy is, in effect, analogous to stray light in a conventional

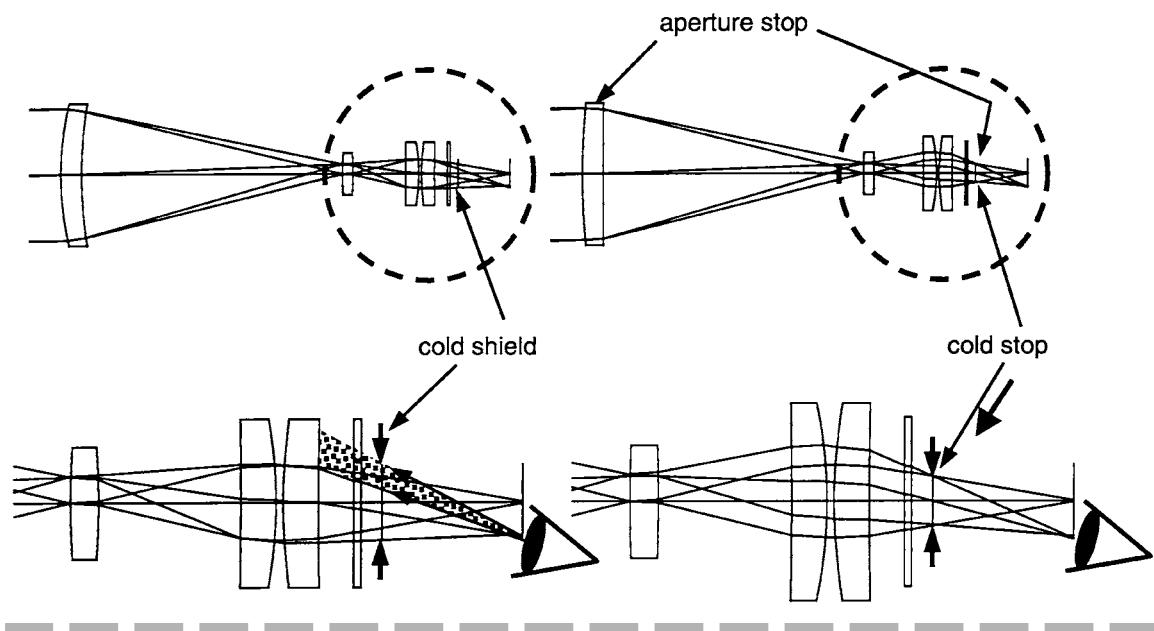


Figure 12.4

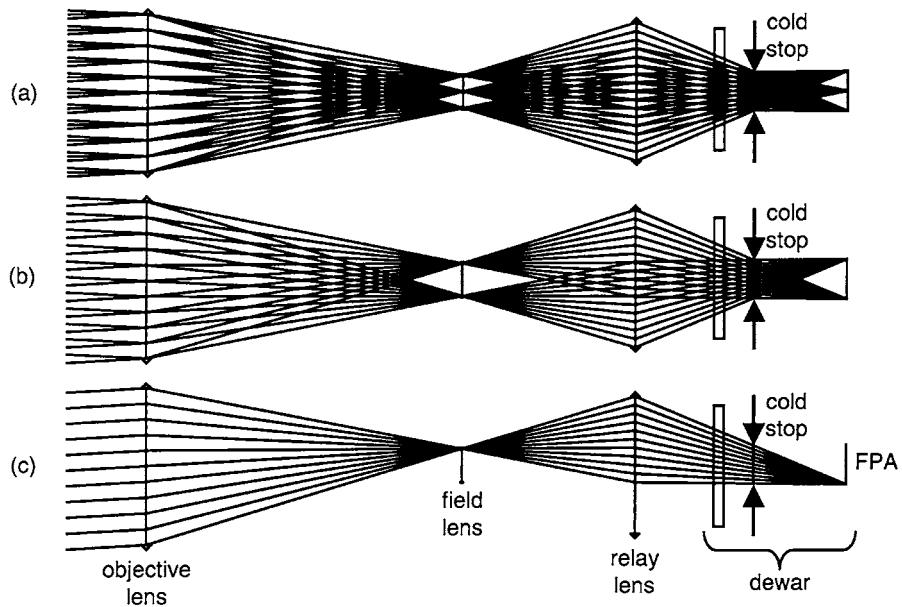
Left: System Which Is Not 100% Cold Stop Efficient; Right: System Which Is 100% Cold Stop Efficient

optical system operating in the visible portion of the spectrum. If this nonscene energy is “warm,” then the sensitivity of the detector will be reduced from its nominal value; however, if this nonscene energy changes in magnitude over the FPA or through scan, then you will have image anomalies similar to ghost images in a conventional visible system. This is bad, especially when the image anomalies are confused with the real scene which can sometimes happen.

On the right-hand figure in Fig. 12.4, we see how the aperture stop at the front element is reimaged into the cold stop plane inside the dewar assembly. Here if we look out from anywhere on the FPA, we see only scene energy and no system interior energy whatsoever. This system is said to be 100% cold stop efficient. The *cold stop efficiency* is the ratio of the total solid angle reaching a given pixel which comes from the scene to the total solid angle reaching the same pixel from the entire opening in the thermal baffle or cold shield. For example, we might find that the solid angle reaching a given pixel from the scene is 90% of the total solid angle which can possibly reach the same pixel within the thermal baffle. This system has 90% cold stop efficiency. The photons or thermal energy which come from some interior portion of the system or its housing is unwanted energy in the form of stray light. Its presence lowers the net sensitivity of the system, but even more importantly if the magnitude of this nonscene energy changes or modulates over the field of view, or through scan if we have a scanning system, then we likely will see image anomalies on the video display.

Figure 12.5 shows a paraxial IR lens system that is designed to be 100% cold stop efficient. In order to make it clear just how the ray paths proceed through the system, we show three permutations of the design. Note that the design itself is identical in all three cases. Figure 12.5a shows three field angles, on axis and $\pm 3^\circ$; Fig. 12.5b shows only the extreme fields of view; and Fig. 12.5c shows only the maximum field of view. Note that in all three cases the front objective lens is reimaged into the cold stop plane. Rays from various fields of view are shown all superimposed. It is clear how the ray cones all overlay on the front element and also on the cold stop surface. It should also be clear how, if you locate your eye anywhere along the FPA and look toward the front of the system, the only solid angle you can see is from the solid angle of image-forming energy. Anything outside these solid angles will view the aft side of the cold stop and/or the interior of the cryogenically cold dewar.

Figure 12.5
Paraxial IR Lens with
100% Cold Stop
Efficiency



Scanning Methods

A typical imaging sensor, such as a CCD or a CMOS chip in a commercial camcorder or other video-imaging system, uses a full two-dimensional detector array or focal plane array (FPA). Very much analogous to film, these FPAs record the entire image essentially at once. In the thermal infrared these FPAs are called staring, mosaic, or two-dimensional (2-D) detectors.

Infrared detectors are still quite costly and difficult to manufacture and for this reason detector arrays which are much smaller in extent than a full two-dimensional array are often used with appropriate scanning to allow imaging coverage over the full desired two-dimensional field of view. If we use a small detector array, we can create a full two-dimensional field of view by following the steps outlined here and illustrated in Figure 12.6.

- Scan the field of view in the azimuth or horizontal direction over the full width of the field. This is the upper swath shown in Fig. 12.6.
- Then simultaneously increment the field down by the vertical extent of the array while reinitializing to the original azimuth position on the left.



Figure 12.6
Serial Scanning

serial scan

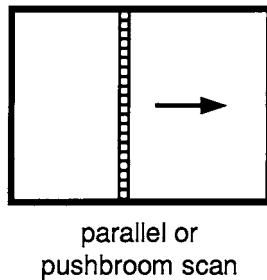
- Now we scan again in azimuth.
- This procedure is repeated until we cover the full vertical or elevation field of view, after which the entire process is repeated again and again to create the full two-dimensional field of view.

This process is known as *serial scanning*. It requires two scan motions, one for the azimuth or horizontal scan and another for the elevation or vertical scan. A good analogy to serial scanning is mowing your lawn. If you have a large rectangular lawn area, you likely use a lawn mower with a width of 0.5 m or so. And you likely mow in one direction, increment by the width of the mower at the end, and mow back. This is known as a *bidirectional scan* in optical terms. Needless to say, you can mow a large area lawn with a small lawnmower, just like covering a large field of view with a small detector array. The previous methodology can be used with very small arrays with as few as one element or pixel or with larger arrays, with up to 25% of the vertical field of view or more. Electronics called *scan-converting electronics* is used to combine or multiplex the data streams and create a virtually seamless standard video signal such as composite video, NTSC, etc.

Operationally, serial scanning requires two scan motions as discussed previously, horizontal and vertical. This can be implemented using two scanning or moving mirrors, one for the scan in the horizontal direction and the other in the orthogonal vertical direction. While, theoretically, both functions could be accomplished with one mirror scanning in both directions, this is rarely done due to the high bandwidths and other difficulties of the two required motions. And do keep in mind that the scan motion in the scan direction should be linear in its angular velocity.

A second form of scanning is known as *parallel* or *pushbroom* scanning, as shown in Fig. 12.7. Here we use a long detector array covering the entire vertical field of view, and the scanning motion is in the

Figure 12.7
Parallel or Push
broom Scanning



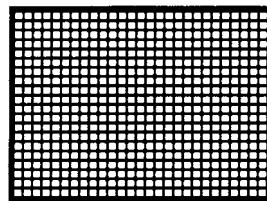
azimuth or horizontal direction (or conversely). For this form of scanning only one scan mirror is required. Comparing this with a serial scan system shows the parallel or pushbroom system requires one scan mirror motion and uses an FPA with more elements. The serial scan system uses two scan mirrors and has more complex scan-converting electronics but a far simpler detector. It is clear that there are many trade-offs to be made in selecting the optimum configuration for a given application.

It is important to note that while we are describing scanning over the field of view in the azimuth and elevation directions, what really happens in the optical systems is quite different. What literally happens is that the optical system is, in reality, quite stationary in space during this scanning operation (except of course for the scan mirror itself). The motion of the scan mirror or mirrors scans or translates the image of object space over or across the image plane or FPA. So if your eyes could see in the thermal infrared and you were able to view the focal plane area, what you would see is the image being translated or scanned across the FPA in azimuth and elevation (unless it is a parallel scan operation in which case the image is moved only in the one direction).

Staring or mosaic arrays are full two-dimensional focal plane arrays which sense the entire scene instantaneously with no mechanical scanning required. Figure 12.8 shows such an FPA. Here we are trading off a more complex sensor for simplified mechanics and scan-converting electronics. It is interesting to note that every commercial camcorder and digital camera uses staring arrays without scanning. The analogous form of sensors are, of course, far more costly in the thermal IR.

We will now show how we can implement a scanner in an IR system for a pushbroom scan. Consider Fig. 12.9 where we show what is, in effect, an astronomical telescope. The system takes collimated radiation of diameter D covering a field of view of $\pm\alpha$ and outputs collimated radiation of diameter d covering a full field of view of $\pm\theta$. For this system the

Figure 12.8
Staring Array (No Scanning Required)



magnification, M , can be stated as D/d or equivalently θ/α . In order to scan over the $\pm\alpha$ full field of view, we could do either of the following:

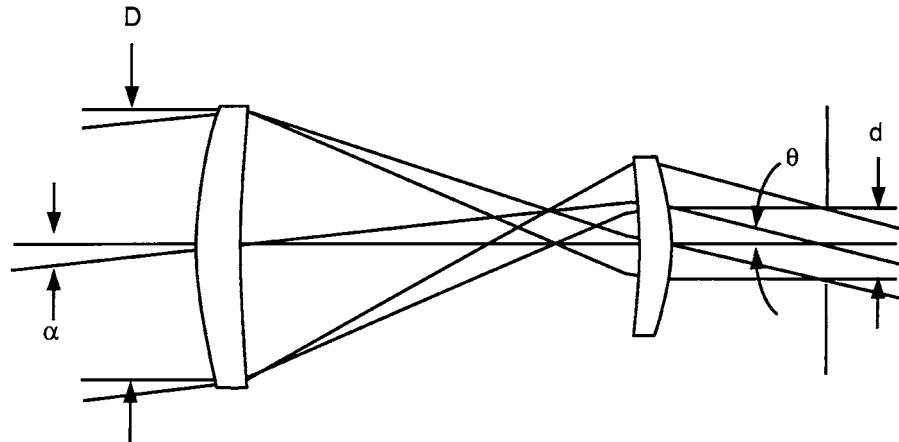
- Place a large flat mirror in front of the system and scan it by $\pm\alpha/2$.
- Place a small flat mirror aft of the second collimating lens at the exit pupil and scan it by $\pm\theta/2 = \pm\alpha(D/d)/2 = \pm\alpha M/2$, where M is the magnification.

Consider the following example: Assume D is 100 mm in diameter and the magnification is $10\times$. Thus, $d = 10$ mm. And assume that $\alpha = \pm 2^\circ$, which means that $\theta = \pm 20^\circ$. We can now accomplish our scanning with either of the following scenarios:

- We can use a mirror approximately 100×140 mm and scan it by $\pm 1^\circ$.
- We can use a mirror approximately 10×14 mm and scan it by $\pm 10^\circ$.

While both of these methods will work, there are many reasons why scanning the smaller mirror by the larger angle is preferred. The smaller

Figure 12.9
Scanning Methodology



mirror is much easier to manufacture. If we require a given surface flatness on the mirror, this requirement will hold for the larger mirror as well as the smaller mirror, and the smaller mirror will be far less costly. Also, the smaller mirror system will be far better from a packaging point of view.

We show in Fig. 12.10 an oscillating mirror for a pushbroom scanning system as described previously. The scanning mirror is located at the exit pupil of the telescope as we described earlier. In Fig. 12.11 we show how a polygon mirror can also be used for a pushbroom scanner. A polygon mirror is a very interesting device. Each facet is, in effect, a separate scan mirror. And each facet, as it scans, rotates not about the vertex of the facet but rather about the center of rotation of the entire polygon. This creates a different motion in space and must be properly modeled to assure proper system performance. Note in Fig. 12.11 that we show the mirror in its “neutral” position as the top mirror. Once again, let’s put our eye at the FPA and look outward. We reflect off of the prime facet to the left and ultimately view object space. Everything is fine. Now look at the bottom polygon mirror, which is shaded. The polygon has rotated around in a counterclockwise direction in the order of 20° , and we show it in a position where the next facet behind the prime facet is just allowing a sliver

Figure 12.10
Oscillating Mirror for
Pushbroom Scanning

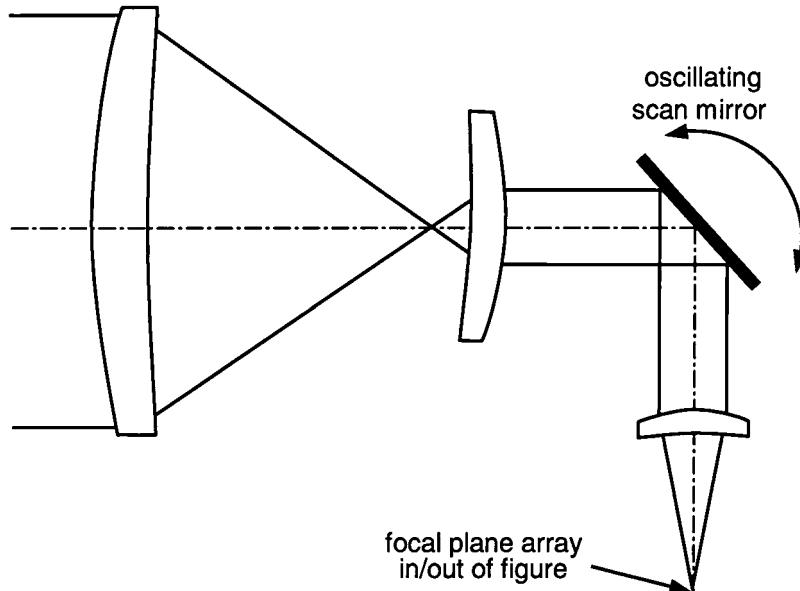
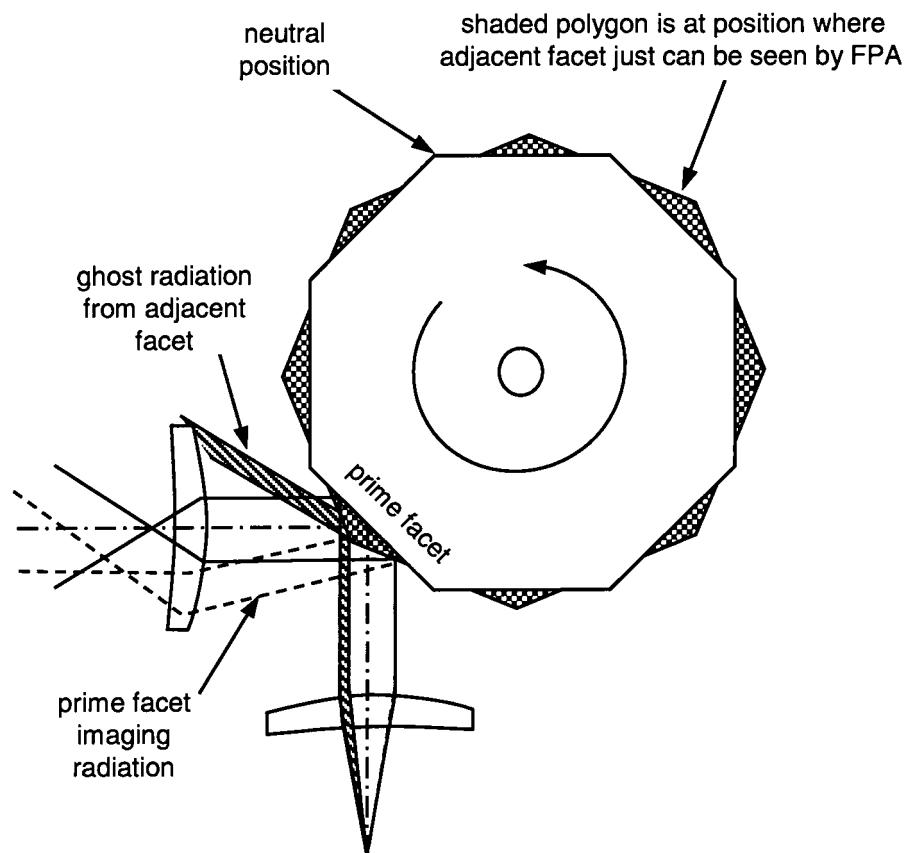


Figure 12.11
Polygon Mirror for
Pushbroom Scanning

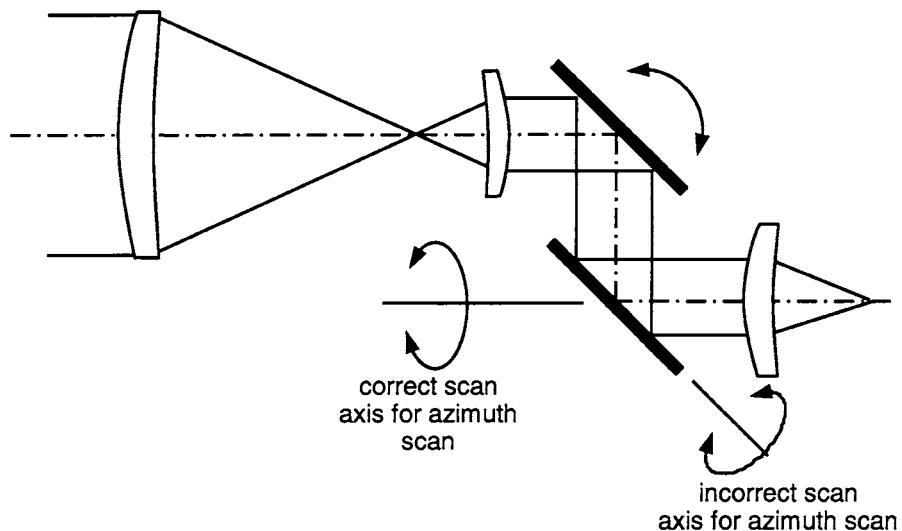


of radiation to be reflected from it. With the polygon in this position, the prime imaging facet reflects the light down to the left while the adjacent facet reflects the sliver of light up to the left. In a situation as shown, the angle of the radiation from the adjacent facet is likely larger than that from the prime facet, and for this reason the radiation may or may not actually reach object space. If it does not, it will strike an interior structure of the system and may create an image anomaly. This reflection from the adjacent facet is known as *ghosting*, and if the ghost radiation makes it to object space it is called *external ghosting* while if it strikes an interior portion of the system it is *internal ghosting*. In order to prevent ghosting, the electronics is often shut off just prior to the adjacent facet entering into the imaging beam of radiation. Occasionally, one can build a valid case to allow a small fraction of the pupil to be ghost

radiation, but it really does need to be small. The fraction of the time that the polygon is actually being used for imaging relative to the total amount of time it is running is called the *scan efficiency*. Scan efficiencies in the vicinity of 80% are not uncommon.

Figure 12.12 shows how we can create a serial scan motion. Recall that in a serial scanning system we scan in the azimuth direction while imaging onto a small detector array of only a few elements. We then increment in elevation and scan in azimuth once again. The process is repeated until we build up a full two-dimensional field of view. Such a system requires two scan-mirror motions, one for the azimuth scan and the other for the elevation direction. In Fig. 12.12 we show the upper mirror as the elevation mirror. For the azimuth scanning mirror we show two potentially viable mirror motions, one is incorrect and the other correct. As it turns out, if we rotate the mirror about the incorrect axis of rotation, we will be scanning an arc in the azimuth direction as opposed to a straight horizontal direction, which will result from the correct axis. It is extremely important that you fully understand just how the mirrors in your system work and how they scan in object space. How can we be sure that we do not have a problem? There are several ways. First, you really should have an accurate model of your scanning system in your computer lens design program. Take nothing for granted! You must spend the time to assure that your model is an accurate representation of the real world and show

Figure 12.12
Two Mirrors (Azimuth and Elevation) for Serial Scanning



how the scan motion works. One interesting thought is to set up your system in reverse from the FPA out toward object space. If you do this, you need only trace a central ray along the optical axis rather than a full cone of rays, as this ray will, in effect, be the chief ray at the given field and/or scan position. Now rotate your mirror or mirrors and monitor at what azimuth and elevation the ray leaves the system for object space. This will tell you directly whether you are scanning a straight line or not. One final hint: take some very simple “drugstore” mirrors and make a crude setup in the lab to model the mirror motions. If in such a setup you exaggerate the mirror motions, you can generally get a valid indication of what is going on. Regardless of how you approach the situation, we cannot overemphasize the importance of modeling properly your system with respect to its scanning motions and the resulting imagery into object space.

IR Materials

While there are many glass types available for visible systems, there are only a very limited number of materials that can be effectively used in the MWIR and LWIR spectral bands. Table 12.1 shows the more common materials and their most important properties. Figure 12.13 shows a plot of the transmittance of the more common IR transmitting materials. It is important to note that these data include surface reflection losses, and often a significantly higher transmittance results after applying high-efficiency antireflection coatings.

Figure 12.14 shows a “glass” map where we plot the refractive index in the ordinate versus the $V\#$ in the abscissa for common infrared-transmitting materials. Recall that the $V\#$ is inversely proportional to the material’s dispersion, and note how for germanium is nearly 1000 in the LWIR (very low dispersion) versus about 100 in the MWIR. You can use this glass map in much the same way as you would for visible systems.

We will now discuss each of the common IR materials:

GERMANIUM *Germanium* is perhaps the most common of infrared materials. It is used in both the LWIR where it is the crown or positive component of an achromatic doublet and in the MWIR where it is the flint or negative component of an achromatic doublet. This anomaly is due to the differences in its dispersion properties in the two spectral

TABLE 12.1

Properties of Common Optical Materials in the Thermal Infrared

Material	Refractive	Refractive	$dn/dt/^\circ\text{C}$	Comments
	Index at 4 μm	Index at 10 μm		
Germanium	4.0243	4.0032	0.000396	Expensive, large dn/dt
Silicon	3.4255	3.4179*	0.000150	Large dn/dt
Zinc sulfide, CVD	2.2520	2.2005	0.0000433	
Zinc selenide, CVD	2.4331	2.4065	0.000060	Expensive, very low absorption
AMTIR I (Ge/As/Se:33/12/55)	2.5141	2.4976	0.000072	
Magnesium fluoride	1.3526	†	0.000020	Low cost, no ctg required
Sapphire	1.6753	†	0.000010	Very hard, low emissivity at high temperature
Arsenic trisulfide	2.4112	2.3816	‡	
Calcium fluoride	1.4097	†	0.000011	
Barium fluoride	1.4580	§	-0.000016	

*Not recommended.

†Does not transmit.

‡Not available.

§Transmits up to 10 μm but drops abruptly.

bands. In the MWIR germanium is approaching its lower absorption band and hence the refractive index is more rapidly changing, thus leading to a greater dispersion. This, in turn, makes it appropriate for the negatively powered element of an achromatic doublet.

With respect to germanium's optical properties, two parameters are of major significance. First, the refractive index of germanium is just over 4.0, which means that shallow curves (long radii) are feasible. As we will see later, along with the higher refractive index aberrations are easier to reduce, which is a significant benefit to the designer. Another parameter of significance is the dn/dt , which is the change in refractive index with respect to temperature. For germanium the dn/dt is $0.000396/^\circ\text{C}$. This is a



Figure 12.13
Spectral Transmittance of IR Materials,
Including Surface Losses

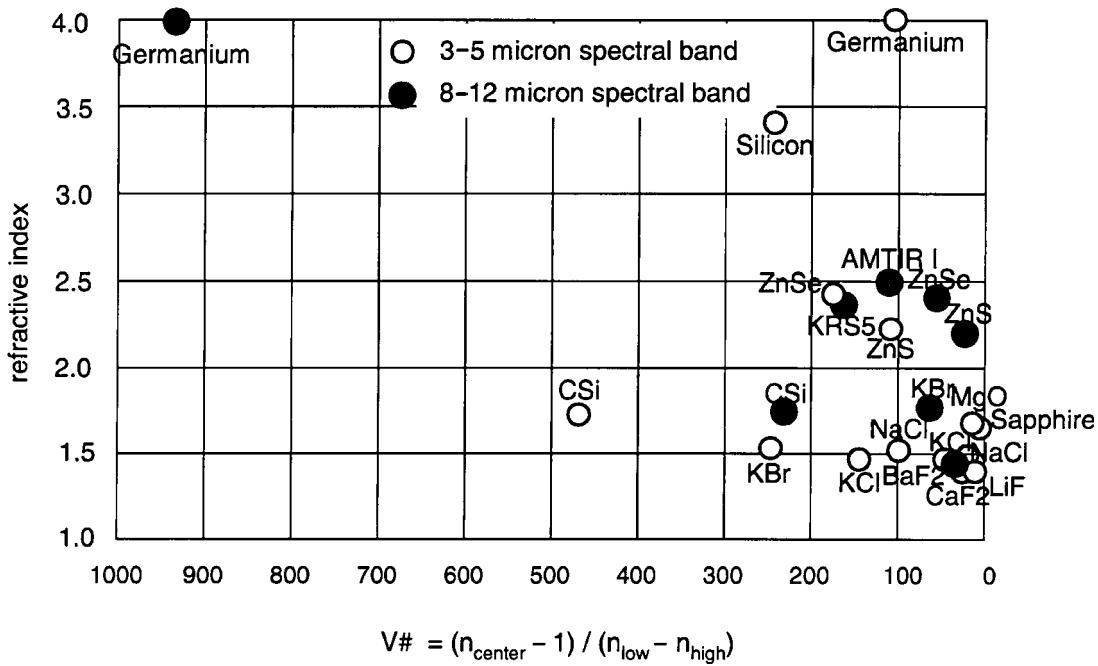
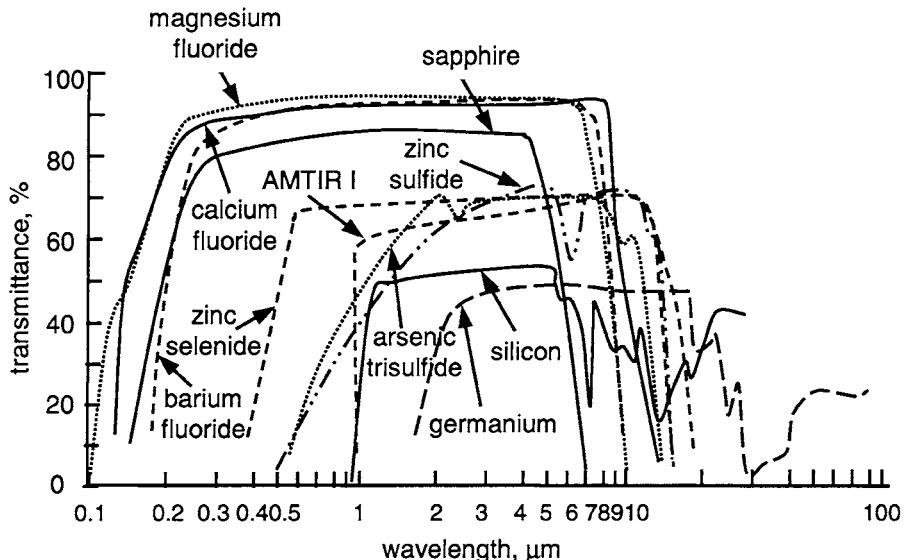



Figure 12.14
Glass Map for Common Materials in the MWIR and LWIR Spectral Bands

very large value, especially when compared with $0.00000360/\text{°C}$ for ordinary glasses such as BK7 glass. This can cause a large focus shift as a function of temperature. This defocus is generally so large that these systems often require some form of athermalization (focus compensation versus temperature).

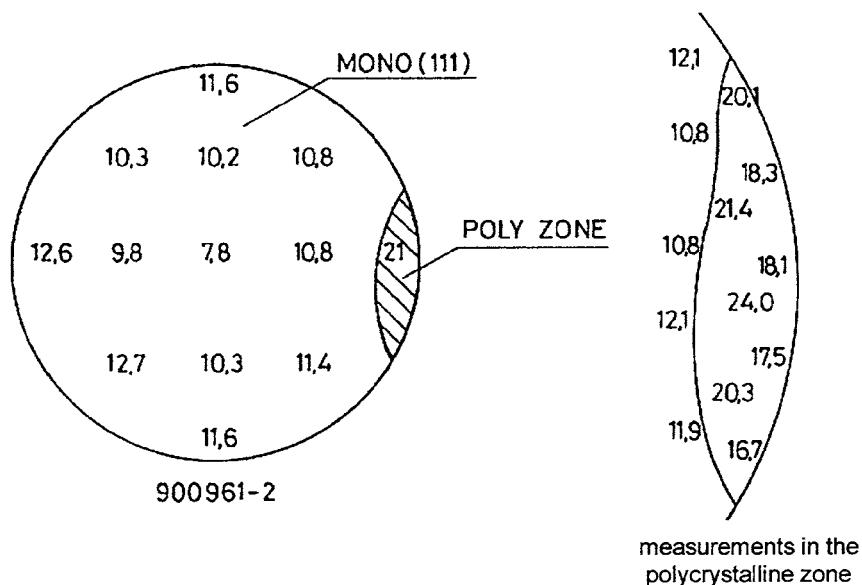
Germanium is a crystalline material and, as such, can be grown in either polycrystalline form or monocrystalline form, which is also called single-crystal germanium. Depending on the manufacturing and refining processes, single-crystal germanium may be more costly than polycrystalline germanium. Throughout the 1970s and 1980s there was much confusion regarding the relative need for single-crystal germanium in high-performance thermal-imaging systems. For the most part, European designers specified single-crystal material, and in the United States polycrystalline material was generally called for. Studies in the mid- to late 1980s showed that indeed polycrystalline germanium had a larger refractive index inhomogeneity, and this was due primarily to impurities at the grain boundaries. Furthermore, the presence of these impurities could be imaged onto the FPA if the material were at or near an intermediate image plane. The single-crystal germanium is preferred. Fortunately, recent advances in material manufacturing have closed the cost differential gap, and, for the most part, the optics industry uses single-crystal germanium. At high temperatures germanium becomes absorptive, with near zero transmittance at 200°C .

Single-crystalline germanium has a refractive index inhomogeneity of 0.00005 to 0.0001, whereas polycrystalline germanium is 0.0001 to 0.00015. For optical purposes germanium is generally specified as to its resistivity in ohm-centimeters, and the generally accepted value is 5 to 40 $\Omega \cdot \text{cm}$ throughout the blank. Figure 12.15 shows a typical germanium blank with an area on the right which is polycrystalline. Note that the resistivity is well behaved and slowly changing radially in the single-crystal region, whereas it changes rapidly in the polycrystalline region. If you were to look into the material with a suitable infrared camera, you would see a somewhat bizarre convoluted image resembling cobwebs, with this appearance most accentuated at the grain boundaries. This is all due to the impurities induced at these boundaries.

There is one further comment regarding germanium—its susceptibility to chipping. You must be exceptionally careful during the optical manufacturing and coating processes as well as during assembly as a nearly inconsequential tap to the edge of a germanium element could result in a chip flaking off. For this reason, germanium is often bonded into its

Figure 12.15

Resistivity Map, in Ohm-Centimeters, for a Polycrystalline Germanium Blank



housing using a semicompliant bonding material. Silicon and some of the other crystalline materials also have this problem, so be very careful.

SILICON *Silicon* is also a crystalline material much like germanium. It is used primarily in the 3 to 5 MWIR spectral band as there is absorption in the 8- to 14- μm LWIR spectral band. While the refractive index of silicon is somewhat lower than germanium (silicon is 3.4255 versus germanium which is 4.0243), it is still large enough to be advantageous with respect to aberration control. Further, the dispersion of silicon is still relatively low. Silicon can be diamond turned with great difficulty.

ZINC SULFIDE *Zinc sulfide* is a common material used in both the MWIR and the LWIR. While its visible appearance varies greatly, it is generally rust-yellow in color and translucent in the visible. The most common process for manufacturing zinc sulfide is known as chemical vapor deposition (CVD).

If zinc sulfide is “HIP’ed” (hot isostatic pressed), it can be made to be water-clear in the visible. While available from several manufacturers, Cleartran is the most common commercially available clear zinc sulfide. Cleartran can be used for multispectral windows and lenses from the visible through the LWIR.

ZINC SELENIDE *Zinc selenide* is in many ways similar to zinc sulfide. It has a slightly higher refractive index than zinc sulfide and is structurally weaker. Because of this, a thin layer of zinc sulfide is sometimes deposited onto a thicker zinc selenide substrate for environmental durability reasons. Perhaps the most significant advantage of zinc selenide over zinc sulfide is that it has a significantly lower absorption coefficient than zinc sulfide. For this reason zinc selenide is commonly used in high-energy CO₂ laser systems.

AMTIR I AND AMTIR III *AMTIR I* and *AMTIR III* are glassy materials manufactured of germanium, arsenic, and selenium in a ratio of approximately 33:12:55. The AMTIR family of materials begins transmitting in the near IR (NIR). For this reason you can often see a very deep and faint red color transmitted through AMTIR. The dn/dt of AMTIR I is about 25% that of germanium, making it attractive from a thermal defocusing standpoint.

MAGNESIUM FLUORIDE *Magnesium fluoride* is another crystalline material. In its crystalline form it transmits from the UV through the MWIR spectral bands. Magnesium fluoride is manufactured by either crystal growth or alternately by “hot pressing.” In this latter process a fine powder form of the material is subjected to very high temperature and pressure in a way similar to powdered metal technology. The result is a milky looking glassy material which transmits well in the MWIR. The caution, however, is that there may be undesirable scattering which could cause contrast degradation and off-axis stray-light problems (this can be avoided using crystalline-grown material). Fortunately, small particle scattering is inversely proportional to the fourth power of the wavelength, so the milky appearance in the visible is reduced by approximately 2⁴ power, or 16 times at 5 μm.

SAPPHIRE *Sapphire* is an extremely hard material (it has a 2000 Knoop hardness value as compared to 7000 for diamond). It transmits from the deep UV through the MWIR. One of the unique aspects of sapphire is that it has a very low thermal emissivity at high temperature. What this means is that the bulk material when at a high temperature will emit less thermal radiation than other materials. You might, for example, use sapphire for the window of a chamber which is subject to very high temperature, especially if you are viewing through the window in the IR. Another application for sapphire is for protective windows of supersonic

vehicles where window heating is a serious problem. The primary disadvantage of sapphire is that its hardness makes it difficult, time consuming, and expensive to optically manufacture. There is a related material whose general class is known as spinel. Spinel is, in effect, analogous to hot-pressed sapphire and can be used in place of sapphire. Spinel is also highly dispersive.

Sapphire is birefringent which means that its refractive index is a function of the plane of incident polarization.

ARSENIC TRISULFIDE *Arsenic trisulfide* is another material sometimes used in the MWIR and LWIR. It is deep red in the visible and is quite expensive.

OTHER VIABLE MATERIALS There are a number of other viable materials including *calcium fluoride*, *barium fluoride*, *sodium fluoride*, *lithium fluoride*, *potassium bromide*, and others. These materials can be used from the deep UV through the MWIR spectral bands. Their dispersion properties make them quite attractive for wide-spectral-band applications, especially from the NIR through the MWIR and on to the LWIR. Many of these materials have some undesirable properties, especially water absorption (they are hygroscopic).

Generally, optical manufacturing methods for IR materials, such as germanium, silicon, zinc sulfide, and zinc selenide, are similar to glass optics. While manufacturers clearly have their trade secrets in this area, to the designer, they can all be considered as manufacturable. Several of the crystalline materials are hygroscopic, which presents some challenges to the optical shop. Also, these materials need to be appropriately coated to prevent damage from moisture and their housings often need to be dry nitrogen purged. IR materials generally have an extremely high index of refraction, which requires antireflection coatings; otherwise the system transmission would be very low.

We have noted the hardness of sapphire previously. One further point of importance is that some IR materials can be single-point diamond turned. These include germanium, silicon, zinc sulfide, zinc selenide, AMTIR, and the fluorides. Sapphire cannot be diamond turned. Silicon can be diamond turned; however, the carbon in the silicon reacts with the carbon in the diamond and this results in a shorter tool lifetime and thus higher cost. Diamond turning can be extremely important if you need either aspheric surfaces and/or diffractive surfaces.

Reduced Aberrations with IR Materials

At the outset of this chapter, we indicated that the higher refractive indices of many of the infrared transmitting materials results in shallower and less steeply curved surfaces which, in turn, results in reduced aberrations. In order to illustrate this, consider Fig. 12.16, where we show six $f/2$ single-element lenses 25.4 mm in diameter, each bent for minimum spherical aberration. The refractive index of the lenses ranges from 1.5 to 4.0, where an index of 1.5 would be close to ordinary BK7 glass and 4.0 close to germanium. Note how the shape of the lenses is progressively changing. At index 1.5, the front is steeply convex and the rear is very slightly convex. At an index of approximately 1.62, the rear surface becomes flat. As the index keeps increasing, the lens becomes more concentric looking. While this is indeed very interesting, unfortunately it does not tell us anything about the aberrations.

Figure 12.17 shows a plot of the rms wavefront error in waves at the wavelength of $0.5 \mu\text{m}$ versus refractive index for lenses bent for minimum spherical aberration. At index 1.5, we have over 10 waves rms

Figure 12.16
Lens Bending As a Function of Refractive Index for Minimum Spherical Aberration

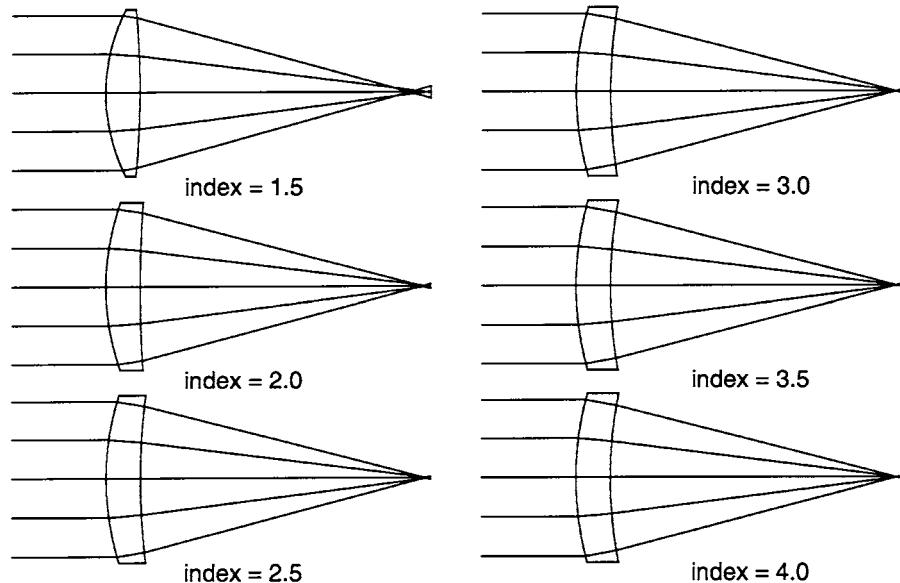
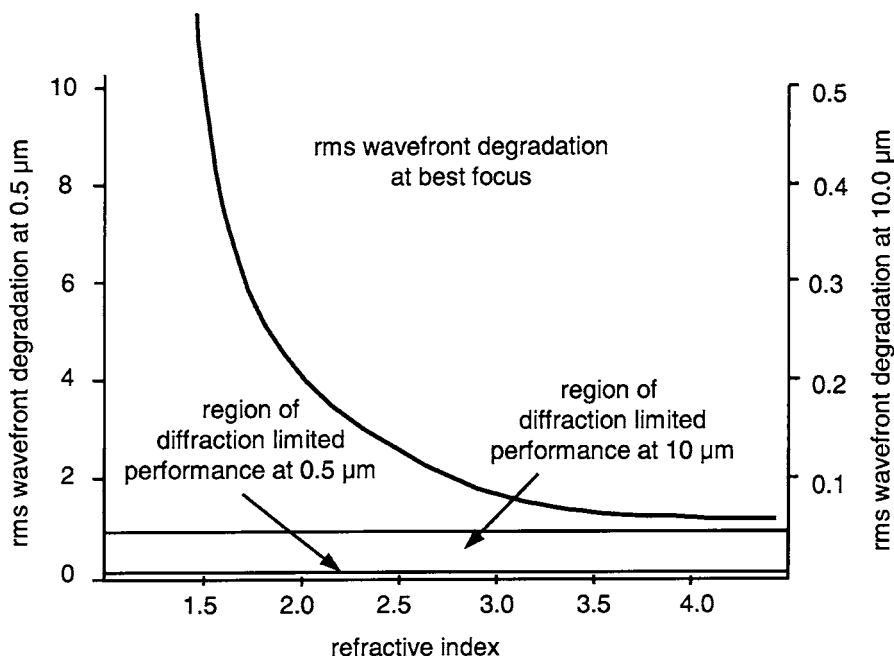


Figure 12.17
Spherical Aberration
As a Function of
Refractive Index for
 $f/2$ Lens 25 mm in
Diameter



which equates to approximately 50 waves peak to valley! This is an enormous amount of spherical aberration. Note how the aberration rapidly decreases with increasing refractive index. At an index of 2.0, which is about as high as we can find for visible glass, we have about 3 waves rms, or approximately 15 waves P-V. Note how at an index of 4.0, we have about 1 wave rms, or about 5 waves P-V. While this decrease in aberration is noteworthy, the most important point is that at an index of 4.0 we really must be thinking of the thermal infrared wavelengths, as glasses are not available with this refractive index. Let's therefore change the scale of the ordinate of the plot to indicate rms wavefront error at 10 μm . Since 10 μm is 20 times the visible wavelength of 0.5 μm , we need to reduce the values in the ordinate by 20 times. Thus, our 1-wave rms becomes 0.05 waves rms, which is approximately 0.25 wave P-V. This now meets the Rayleigh criteria and is diffraction limited! To summarize this extremely noteworthy finding: We have shown that a single $f/2$ element of glass 25.4 mm in diameter which is bent for minimum spherical aberration and has a refractive index of 1.5 similar to BK7 has approximately 50 waves P-V in the visible (200 times the diffraction limit). An equivalent single element of germanium with refractive index 4.0 when bent

for minimum spherical aberration and referenced to a wavelength of $10 \mu\text{m}$ in the LWIR just meets the Rayleigh criteria with approximately 0.25 wave P-V!

It is for this reason that infrared optical designs are generally simple in form as compared with their visible counterparts. While a single germanium element may indeed suffice for an $f/2$ LWIR lens 25.4 mm in diameter, in the visible we would require three separate elements to achieve a diffraction-limited performance.

It was noted earlier that some infrared materials have very low dispersion, and for this reason color correction may not be required in some scenarios. In order to demonstrate this, we have come up with a virtual lab experiment right here in the book. Figure 12.18 shows the setup. We have a prism located 2.5 m from a vertical wall. We will now manufacture prisms similar to what is shown of various materials so that the center of the respective spectral band will be descending at a $(45-\alpha)^\circ$ angle, where α is the prism angle. We will then measure the length of the resulting spectrum for the visible, MWIR, and LWIR spectral bands.

Figure 12.19 shows the results. Highly dispersive SF6 glass used in the visible from 0.4 to 0.7 μm has a large spectrum measuring about 120 mm. Less dispersive BK7 glass would have a spectrum measuring about 30 mm. In the 3- to 5- μm spectral band zinc sulfide would have a spectrum about

Figure 12.18
Hypothetical Experiment to Show
Length of Spectrum
for Different Materials

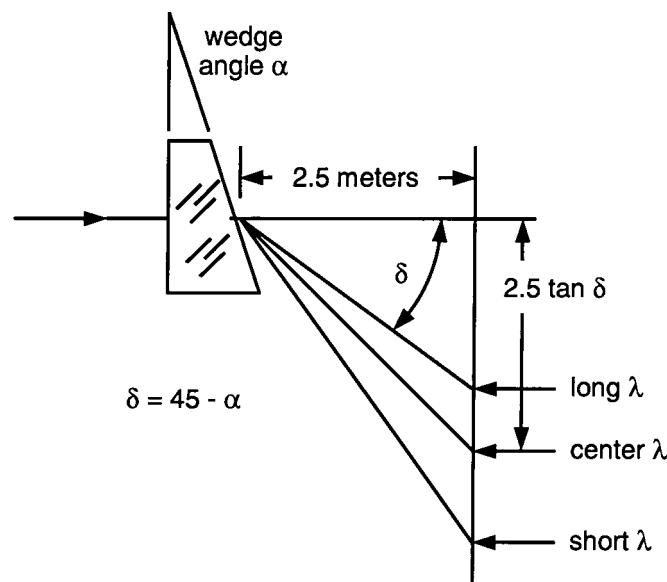
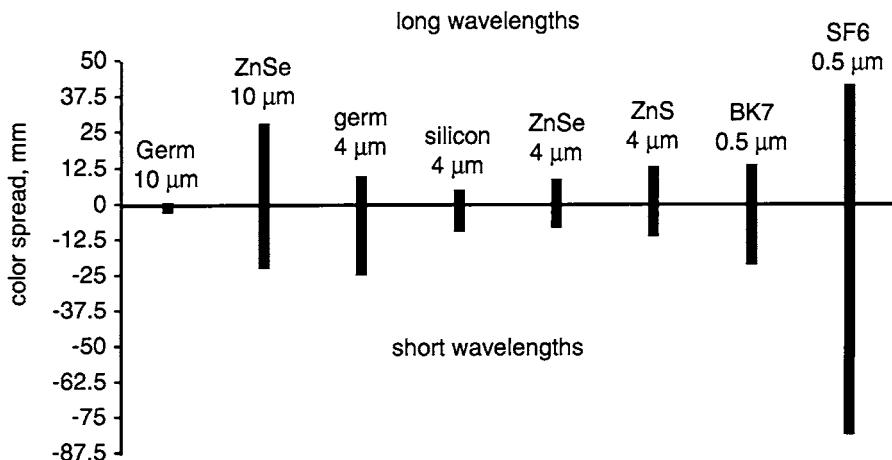


Figure 12.19
Length of Spectrum
on Wall 2.5 m from
Prism



35 mm long, zinc selenide about 17 mm long and silicon about 12 mm. Germanium, as we know, is quite dispersive in the MWIR spectral band, and its spectrum would be about 35 mm. In the LWIR we find that the spectrum for zinc selenide is about 50 mm long and finally germanium in the LWIR is about 4 mm long. This is significantly less than any of the other materials, and this is why we can often use germanium alone in the LWIR spectral band without the need for color correction.

Image Anomalies

Thermal infrared systems often show cosmetically undesirable image anomalies which are not seen in visible optics. These effects include *narcissus*, *scan noise*, *beam wander*, *ghosting*, and *shading*. The effects are similar to what we generally think of as ghost images, and the resulting imagery can vary from slight brightness variations over the format to sharp bright or dark areas. While the mechanisms differ, all of these effects are due to the detector seeing more (or less) thermal energy over the field of view or through scan than dictated by the scene energy itself.

As we discussed in the section “Cold Stop Efficiency” earlier in this chapter, one of the most important methods for evaluating the properties of thermal infrared systems is to put your eye “figuratively” at the detector and look forward (into the exit pupil) and ask yourself “what do

you see." This is sometimes called the *detector's eye view*. For an IR system with 100% cold stop efficiency you should see a solid angle containing only scene energy, and everything outside this solid angle should be cryogenically cold. If this is the case, then you will indeed be accurately recording or imaging the thermal radiance from the scene. However, if you can see any thermal energy outside the solid angle representing scene energy and inside the cold stop solid angle, this represents extra energy, which will behave in a similar fashion to stray light in visible systems. We will now review the primary causes of image anomalies in thermal imaging systems.

NARCISSUS *Narcissus* occurs because of a change in the magnitude of radiation reflected back into the dewar from lens surfaces within the system. Consider Fig. 12.20 where we show a scanning IR system at the center of scan (Fig. 12.20a) and at the end of scan (Fig. 12.20b). The focusing optics and detector are shown as rotated about the scan pivot point in Fig. 12.20b which is an optically valid representation of what is happening. We also show an enlarged view of the aft end of the system in Fig. 12.20c. We will be discussing only what happens from the rear surface of the last lens element. The total Narcissus effect is the radiometric summation from all lens surfaces.

If at the center of scan (Fig. 12.20a) we look out from the center FPA pixel, this ray will travel along the optical axis and reflect right back on itself ultimately returning through the cold stop into the dewar. Thus this ray "sees" only cold radiation from inside the dewar, or perhaps better said, no radiation at all. Now let's consider the ray from the center of the FPA which just passes by the edge of the cold stop. This ray, after reflection from the lens surface, diverges on its way back and misses entirely the cold stop aperture. In fact, this ray will strike some portion of the system interior, which could be ambient temperature or it could be a hot electronics board. The total solid angle, which can return into the cold stop is shown as the shaded solid angle in Fig. 12.20.

Now consider Fig. 12.20b, which represents the edge of scan. Here all of the energy within the entire solid angle within the cold stop reflects from the lens surface into the system interior. None of this energy will return into the cold stop and "see" cryogenically cold temperature. We thus can see how there is more "cold energy" (or lack of warm energy) at the center of scan than the edge of scan. If we now sum up the radiometric effect from each lens surface, and if we take into account the difference in solid angle returning into the dewar at the center and the

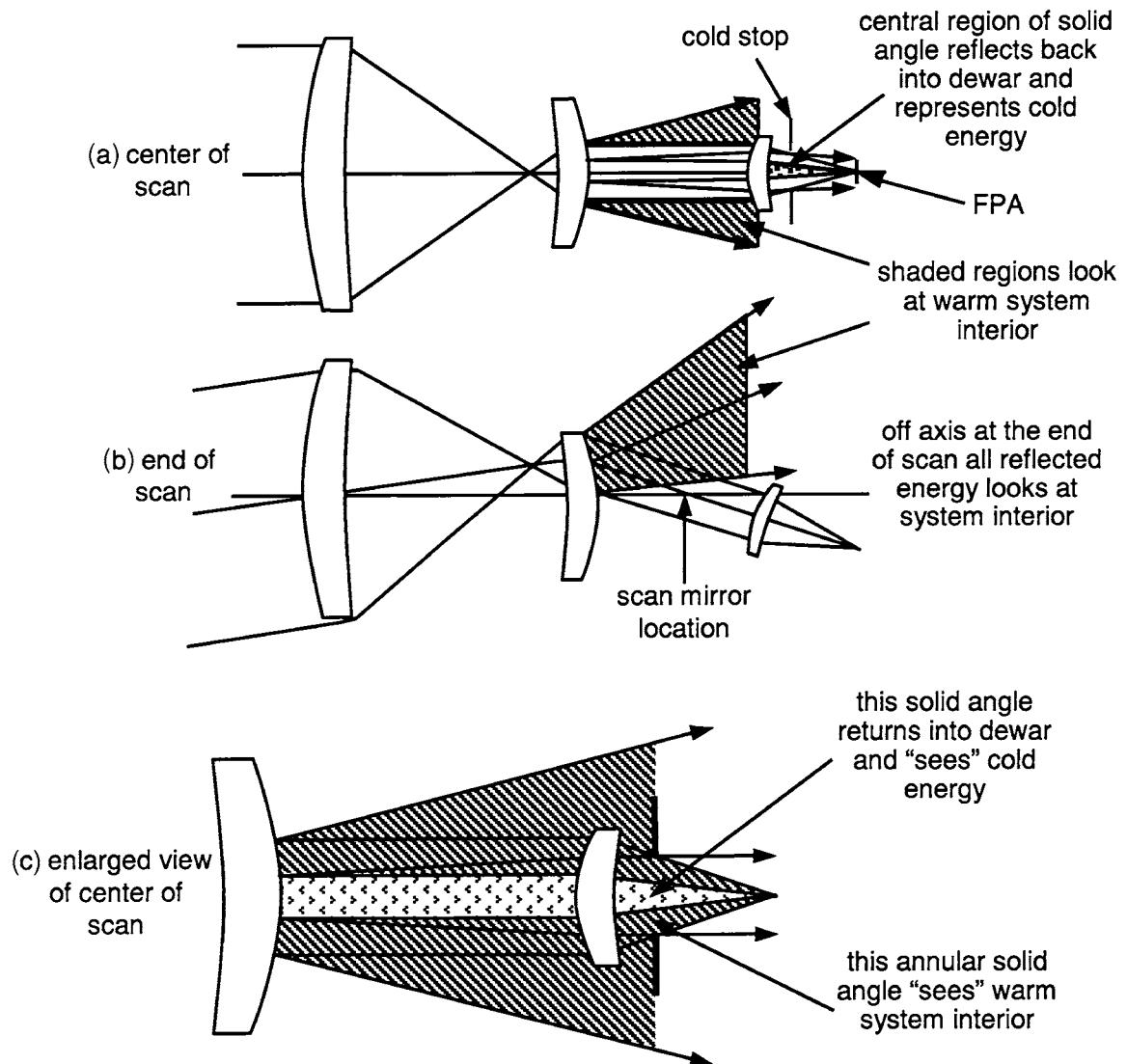


Figure 12.20
Illustration of How Narcissus Is Formed in a Scanning System

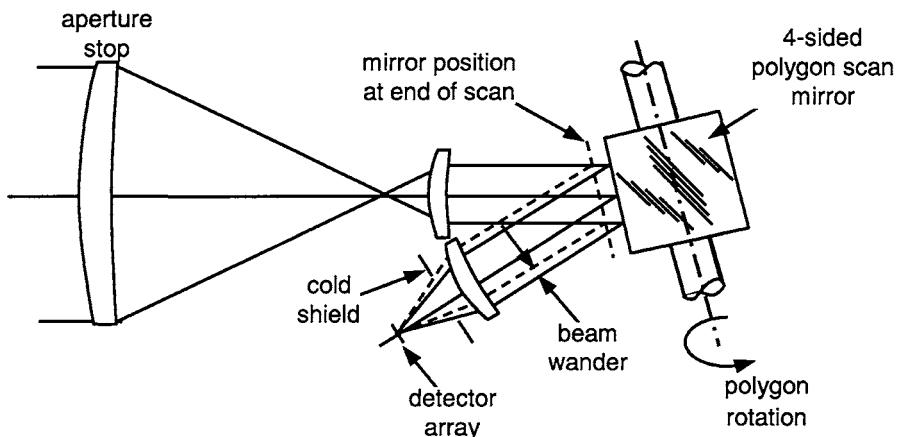
edge of scan, we will find that there is more “warmth” seen at the edge of scan than at the center of scan. This is, of course, due to the summation of the dark solid angles of energy returning into the dewar at the center of scan from each lens surface. The bottom line is that the video monitor will show less thermal energy at the center of scan than the edge of scan, and this may appear as a dark porthole or disk on the monitor. The diameter of the dark disk and the gradient from the center of the monitor to the edge will depend on the specific geometry and how rapidly the cold energy solid angle reduces with scan.

Note that in nonscanning staring array systems using two-dimensional FPAs you can still have Narcissus. There is one important difference in this regard between scanning and nonscanning or staring systems—the practice in staring systems of performing a “nonuniformity correction.” What this means is that the system is periodically aimed toward a uniform thermal source, and then each pixel is adjusted in its offset to yield a constant gray level over the entire format. For many thermal imaging applications where we are looking at near-ambient temperatures, simply draping a black cloth over the front of the system and then performing the nonuniformity correction will suffice. However, if the temperature of the system interior where the reflected radiation strikes at the edges of scan changes, there may be a need to perform a new nonuniformity correction.

How can we minimize the effect of Narcissus? There are two basic methods to minimize Narcissus. First we can use so-called anti-Narcissus coatings. These antireflection coatings typically have from 0.2 to 0.3% average reflectivity from 3 to 5 μm or alternatively 8 to 14 μm . The second approach is to change the relative bendings of the offending lens surfaces so as to minimize the cold solid angle of reflected radiation. This is a common technique, and to keep the desired level of optical performance, often requires the use of aspheric surfaces. Fortunately, this is not a major problem with the use of single-point diamond turning, now a mature technology. Optical systems often have a flat protective window in front of the telescope. In order to avoid the Narcissus generated by the window, the window is tilted, so that the reflected radiation falls out of the sensor’s field of view.

GEOMETRICAL SCAN EFFECTS *Geometrical scan effect (beam wander)* is illustrated in Fig. 12.21, where we show a generic four-sided polygon scanning IR system. We will assume that the system aperture stop is on the front lens element, a likely situation. At the center of scan, the

Figure 12.21
Geometrical Scan
Effects and Beam
Wander



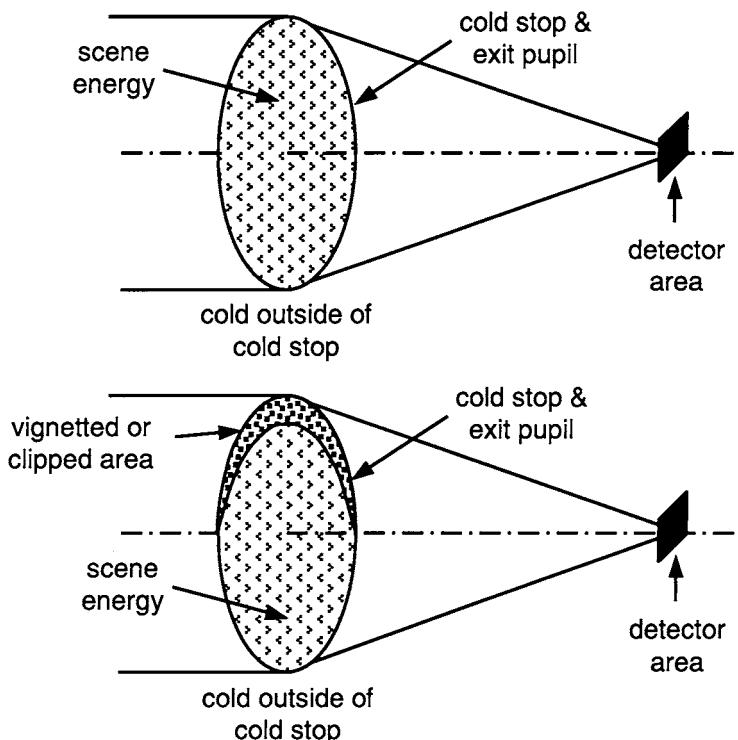
Polygon will be at its nominal or neutral position as shown. As the polygon rotates to the end of scan, its geometrical position in space, where the radiation reflects from its surface, translates to the dashed position (to the left). This represents the location where the intersection of the polygon facet with the plane of the figure has translated for end-of-scan imaging. Since we have stated previously that the aperture stop is on the front element, the beam of radiation reflected from the facet must translate or shift laterally by an amount called the *beam wander*.

We now have a dilemma... should the cold stop be sized for the center of scan and thus clip energy from the beam-wandered edge of scan radiation or should we increase the diameter of the cold shield (along with a lateral translation of the cold shield)? The "lesser of the two evils" is generally to increase and laterally shift the cold shield so as to eliminate clipping of the imaging radiation from any scan position.

SCAN NOISE *Scan noise* is often taken to be any undesirable change in nonscene thermal energy reaching the FPA through scan, thus causing anomalies such as bright bands and other image defects. The appearance of scan noise can often resemble flare and stray light in visible systems. One of the more common causes of scan noise is clipping or vignetting which can result if a portion of the housing infringes on the radiation bundle at the ends of scan.

Consider Fig. 12.22 where we show a representation of the solid angle of radiation reaching the detector at the center of scan (top) and at the edge of scan (bottom). At the center of scan, the solid angle is totally

Figure 12.22
Effect of Clipping or
Vignetting



scene energy within the cold stop and cryogenically cold outside the cold stop. Thus, the only energy the detector will “see” or record is from the scene. At the edge of scan, the solid angle within the cold stop comes primarily from the scene; however, there is a crescent-moon–shaped shaded area which is caused by clipping or vignetting from a housing interior. The solid angle outside the cold stop is cryogenically cold as for the center of scan case. Thus the only difference from the center to the edge of scan is the clipping from the housing interior.

It can be shown that the perceived change in temperature is approximately given by

$$dt \sim \frac{\Delta A}{A} (\text{temperature clipped} - \text{temperature scene})$$

The quantity $\Delta A/A$, is the percent of area of the exit pupil within the cold stop which is clipped. Let us take some representative values for the preceding quantities. Assume that $\Delta A/A = 1\% = 0.01$, the temperature

of the clipped housing is 50°C , and the scene temperature is 0°C . This gives us a perceived temperature difference from the center of scan to the edge of scan of $0.01 \times 50^{\circ}\text{C} = 0.5^{\circ}\text{C}$. If we are scanning in the azimuth direction with a polygon or alternatively with an oscillating mirror, the net result will likely be bright bands at the left and right of the monitor, which get brighter toward the corners of the field of view.

With today's detector technology, temperature differences far below 0.1°C can easily be seen. It is thus clear that even 0.1% clipping can often be seen and may be a problem. In fact, a piece of dirt on a lens element measuring only $0.4\text{ mm} \times 0.4\text{ mm} = 0.16\text{ mm}^2$ is equivalent to 0.1% clipping for a 25.4-mm-diameter ray bundle.

HOSTING *Ghosting* is a term applied to an effect unique to polygon scanners. Consider Fig. 12.11 where we showed radiation incident onto a polygon scanner facet. If the polygon is in its neutral position with the facet tilted at 45° to the incoming radiation, the radiation simply reflects downward toward the FPA. Recall that to best evaluate IR systems relative to image anomalies, you should put your eye at the detector and look out and ask yourself "what do you see?" Let's do this for the facet in the central position. All we see is scene energy, as expected. However, now rotate the facet around in the counterclockwise direction until the next adjacent facet has just entered the beam of radiation. If we look out from the FPA, we will see energy reflected from the prime facet heading down to the left and a small sliver of energy from the adjacent facet heading up to the right. This sliver of radiation will be at a greater angle numerically than from the prime facet, and thus it may miss the lenses and hit an interior portion of the housing. Alternatively, this energy may make it out into object space in which case it is viewing a portion of the scene at the opposite side of the field and outside the nominal field of view. When it is hitting a housing interior, this is known as "internal ghosting" and when it views object space it is known as "external ghosting." Thus, a person holding a bright flare just outside the field of view on the right side may appear inside the field of view on the left due to external ghosting.

What is required to best understand effects such as ghosting is to perform an accurate three-dimensional modeling of your system. Every aspect of the system must be properly modeled, including the polygon axis of rotation, facet clear aperture, and the clear aperture of other lenses within the system. Only then will there be some hope of accurately predicting what is happening.

SHADING *Shading* is very different from any of the previous phenomena, and it is more difficult to explain. Consider the exit pupil diameter at the center of scan reaching the focusing optics. If we have a 100-mm-diameter entrance pupil and a magnification of $20\times$, then the radiation bundle diameter at the exit pupil is 5 mm. Now as we proceed off axis in field of view and scan, what happens to our 5-mm beam diameter? For example, let's assume that we have positive distortion off axis, which means that the magnification increases with field of view. Thus, the 5-mm exit pupil will be smaller in scan mirror space and as it enters the focusing optics. If the aperture stop is at the front element, this will result in some area of the housing being seen at off-axis field and scan positions. The net result is a brightening of the display monitor away from the center of the field of view due to less scene energy and more housing energy.

Athermalization

As any optical system is subjected to higher or lower temperatures several things happen: the lens elements expand or contract, the housing expands or contracts, and the refractive index of the lens materials increases or decreases. While the lens and housing relative expansion can sometimes be a problem, the primary problem with infrared systems is the very large change in refractive index as a function of temperature, or dn/dt . Germanium has a dn/dt of $0.000396/\text{ }^{\circ}\text{C}$. For comparison, BK7 glass has a dn/dt of $3.6 \text{ E-}6$. It can be shown for a simple lens that

$$df = \text{change in focal length} = \frac{f}{n - 1} \frac{dn}{dt} \Delta t$$

where df is the change in focal length and Δt is the change in temperature in degrees Celsius.

Consider the following example: Assume that we have a 75-mm-diameter $f/1.5$ germanium lens with a focal length of 112.5 mm. Applying the previous relationship yields a change in focal length of 0.599 mm for a $40\text{ }^{\circ}\text{C}$ Δt thermal soak. For reference, the Rayleigh criteria for one-quarter wave of defocus is ± 0.046 mm, so the preceding defocus value equates to 3.3 waves of defocus, which is 13.1 Rayleigh criteria depths of focus, a huge amount! This issue can, and often is, a very serious problem in thermal infrared systems. In this example, if we were to control the

temperature so as to stay within a one-quarter wave of defocus, we would need to control the temperature to within $\pm 3^\circ\text{C}$.

It is thus apparent that unless the user can actively refocus the lens, athermalization is imperative. We could translate the FPA or the prime collecting lens; however, this is generally impractical. We could translate another lens element by a greater or lesser amount depending on the magnification, and this is often the approach. Two other approaches are also viable: negative elements with a high $d\text{n}/dt$ can be used and use of reflective optics.

To show how effective reflective optics can be in achieving athermalization, consider Fig. 12.23. We show here two very similar systems, a fully refractive system above the optical axis and a system below the axis where a Cassegrain reflective system takes on the role of the two elements following the curved window of the refractive system. The system has a 75-mm entrance pupil diameter and is used in the LWIR at 10 μm central wavelength. Table 12.2 shows the defocus contributions for the refractive and its reflective counterpart for a 50°C thermal soak (uniform temperature change). First, look at the refractive system as in Fig. 12.23. The front

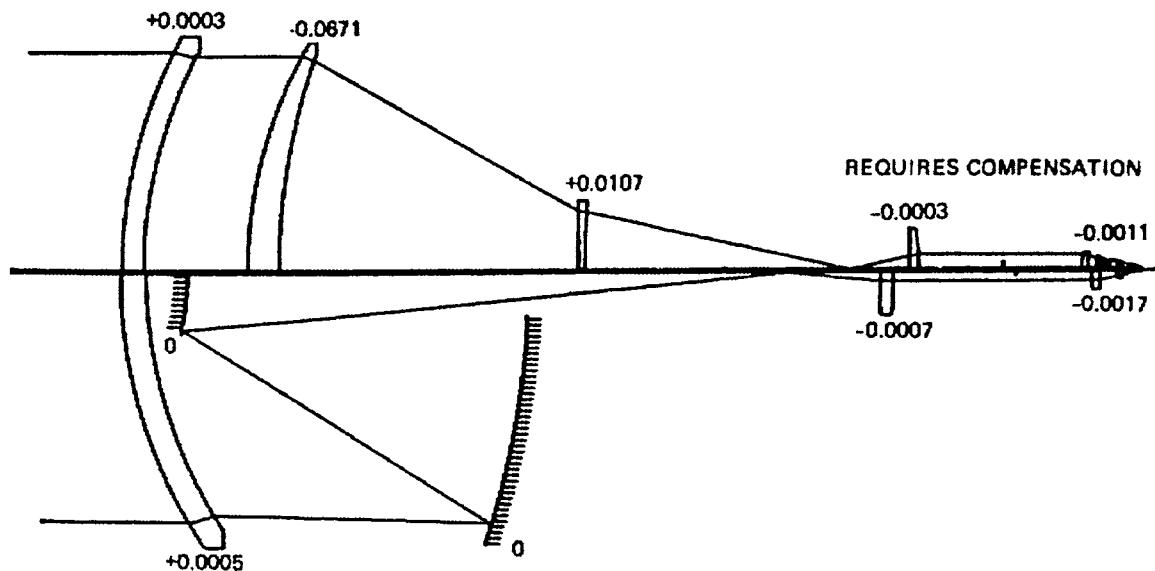


Figure 12.23

Thermal Sensitivities for Refractive System (Top) and Reflective System (Bottom)

TABLE 12.2

Thermally Induced
Focus Shifts for
Refractive and
Reflective IR Optical
Systems

Parameter	Refractive System (mm)	Reflective System (mm)
Curved dome	+0.0076	+0.0127
Large lens or mirror	-1.704	0.0
Next lens or mirror	+0.272	0.0
First small lens (collimating)	+0.0076	-0.018
Second small lens (focusing)	-0.028	-0.028
Entire system	-1.712	-0.048

curved dome has very little optical power (the power is slightly negative) and its contribution is to defocus the image outward 7.6 μm . The large first element accounts for a 1.7-mm inward image defocus and the negative next element moves the image outward by 0.27 mm. The last two elements cause only a small defocus. The net system defocus due to the 50°C thermal soak is about the same as the large powered element at 1.71 mm inward, which is the same as the first element alone.

Looking now at the reflective Cassegrain system, we see that the two mirrors cause no thermal refocusing at all, and the total system defocus is 0.048 mm inward which just happens to be extremely close to the quarter wave Rayleigh criteria.

These data are summarized in Table 12.2.

It is important to understand just why the thermal defocus of the reflective Cassegrain is zero. In fact, there is a subtlety here—each mirror does cause a significant defocus with temperature because it changes shape with temperature. However, the total reflective optics system (the Cassegrain) has virtually zero thermal effect. Perhaps the best way to convince yourself that this is true is to consider a Cassegrain system with only a primary and a secondary mirror along with its support structure, and draw the rays coming from infinity, converging from the primary to the secondary, and finally focusing onto the image plane. Assume that the mirrors and all of the support structure supporting the mirrors and the FPA are all made of the same material such as aluminum. Under a positive thermal soak condition, the entire system will expand uniformly, and the system will therefore stay in perfect focus. To be fully convinced of this conclusion, simply think about taking the drawing you have just made and enlarge it on your copy machine; then look at the enlarged drawing and notice that it is still in focus! So when

we assign a zero sensitivity to the mirrors, we are assuming that this is in the context of a total reflective system of the same material.

We have shown in the preceding example how a nearly all-reflective system is athermalized to within the Rayleigh criteria. How do we restore focus in the refractive system? We could axially translate the detector or alternately translate the front element aft of the curved window. Both of these approaches are unattractive. Moving the detector/dewar assembly is just not a good idea with all of the electronic as well as cryogenic connections, and translating the massive first element is also a difficult task. The best approach is to translate another element within the system by an appropriate amount, which will be dependent on its magnification within the system. In the system shown, the best candidate is to move axially the negative element aft of the large front lens element. To first order, we will need to translate this element about the same amount as the image shifts from the large front element, and this is about ± 1 mm rough order of magnitude. Thus, as the system heats up by 50°C , the image from the large front element moves forward by about 1 mm, which means that the negative element must also move forward by approximately 1 mm in order to restore focus on the final image sensor. The specific magnitudes of motion need to be fully verified using your computer model. If this were the method to be employed, you could locate two to three thermisters or other temperature sensors at the outer periphery of the first element and drive the compensating element (element 2) a distance stored in a look-up table based on the temperature. The temperature sensors would be at this position in the system since the large first element is by far the most sensitive to thermal defocus, as shown earlier. Alternately, an active system approach similar to that used in 35-mm and other camera systems can be used, whereby the algorithm is based on maximizing scene contrast or some other similar criteria is used.

One of the most important points to be made here is that any element motions for athermalization reasons must be sufficiently accurate to locate the image to within approximately one Rayleigh criteria depth of focus from the nominal. In the examination of which element or elements to move, you must examine the total range of motion required. Within this range, you need to determine the finest focus that needs to be made. Make sure that this fine focus adjustment is achievable with the envisioned motion mechanism; otherwise, there will be conditions where a sharp focus cannot be achieved.

System Design Examples

To illustrate how lens designs for the thermal infrared look and work, we show in Figs. 12.24 through 12.31 designs for an $f/2$ LWIR lens 25 mm in diameter covering a field of view of $\pm 2.5^\circ$. The materials include germanium, AMTIR 1, zinc sulfide, and zinc selenide. We show for each design a layout, the transverse ray aberration curves on axis and at 2.5° off axis. Also tabulated is the design prescription data along with the rms wavefront error, which is the average of the axis, 1.25° off axis and 2.5° off axis. With these data you can set up these designs and work with them as you wish. Do note that there is no consideration given to cold stop efficiency in these examples.

The resulting performance is shown in Table 12.3.

In Fig. 12.32 we show an $f/3$ lens with 100% cold stop efficiency along with its design data. Note that this design form is different from that shown in Fig. 12.4 where we achieved 100% cold stop efficiency by reimaging the first element, which also was the aperture stop into the cold stop. Here we are not using any reimaging at all, and in order to make the cold stop in the dewar the aperture stop it was simply set up that way. Notice the extreme beam motion over the front elements and

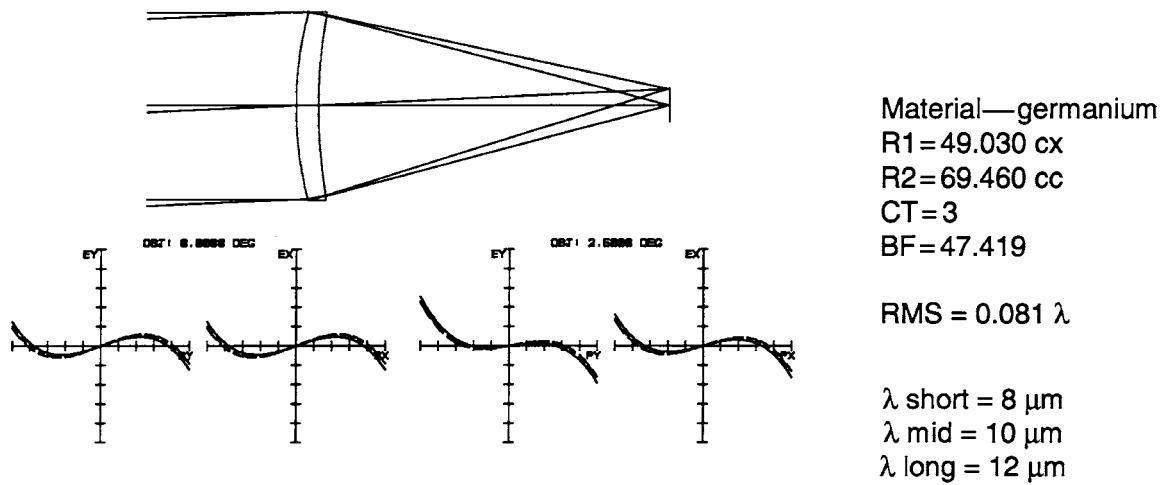
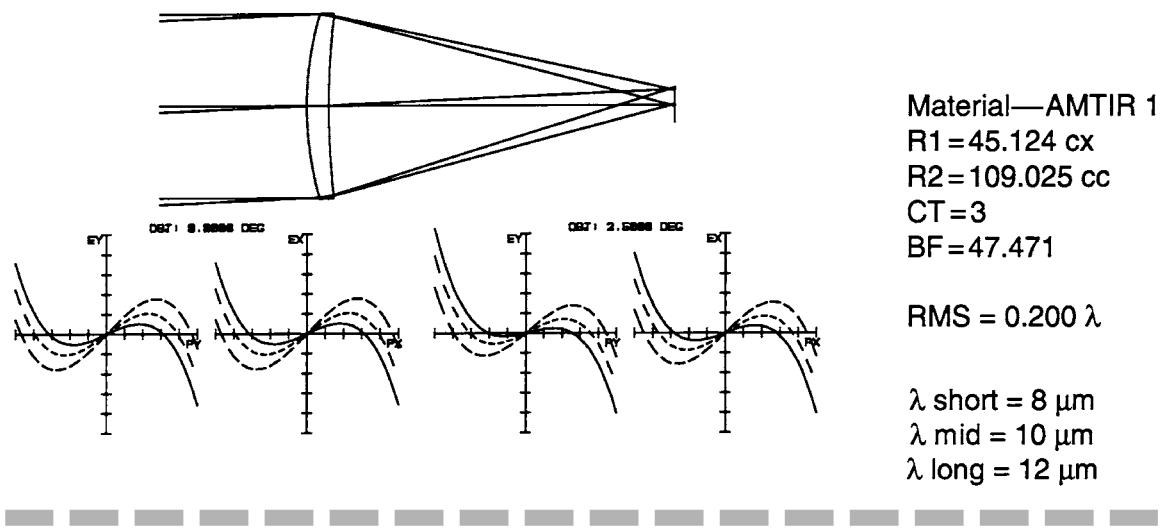
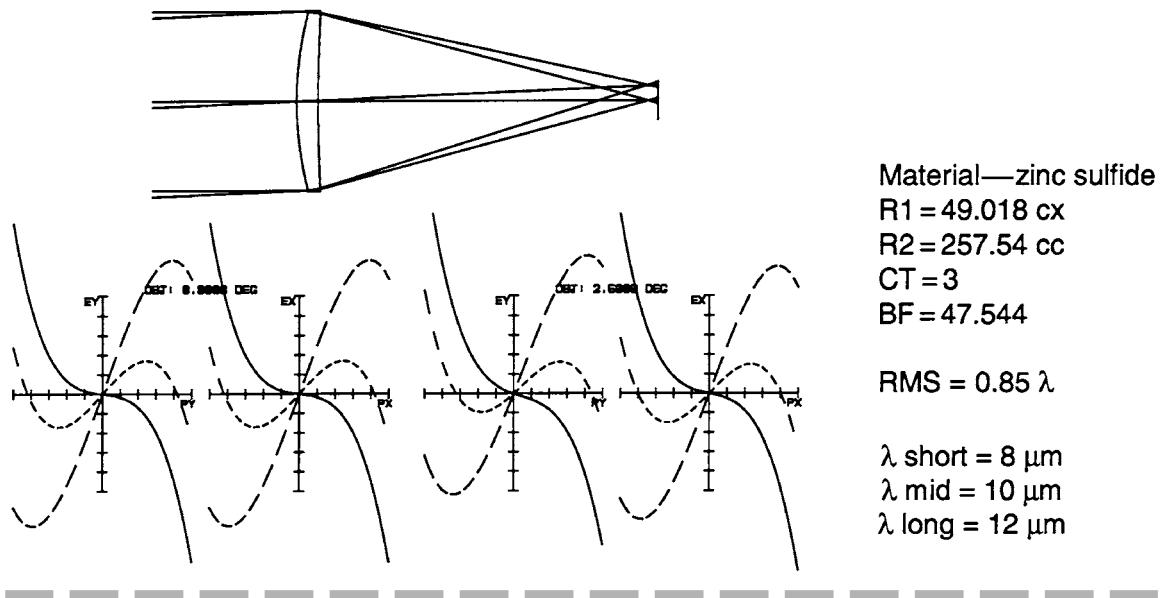


Figure 12.24
Single-Element Germanium Lens, $f/2$ 50-mm Focal Length

**Figure 12.25**Single-Element AMTIR 1 Lens, $f/2$ 50-mm Focal Length**Figure 12.26**Single-Element Zinc Sulfide Lens, $f/2$ 50-mm Focal Length

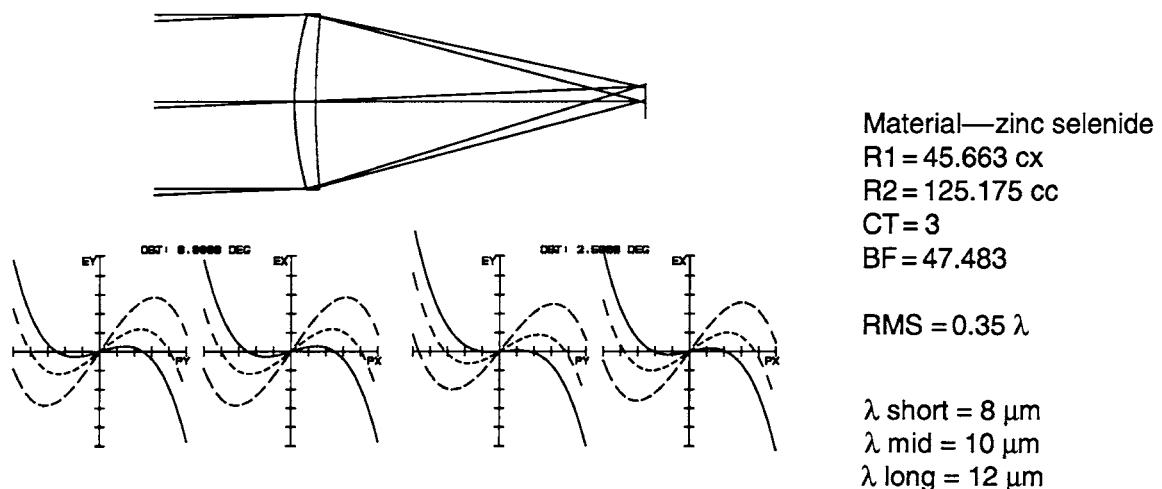


Figure 12.27
Single-Element Zinc Selenide Lens, $f/2$ 50-mm Focal Length

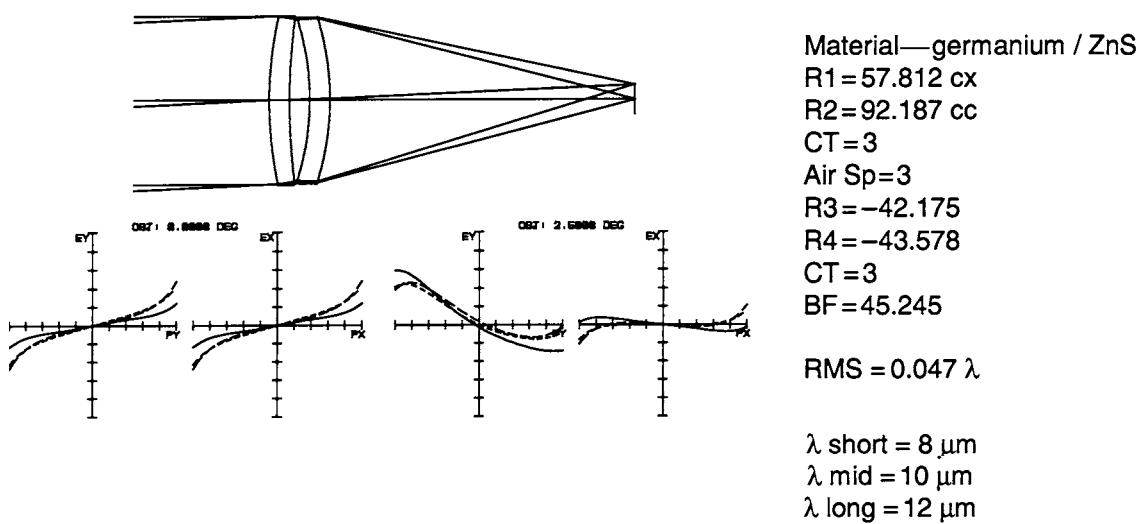
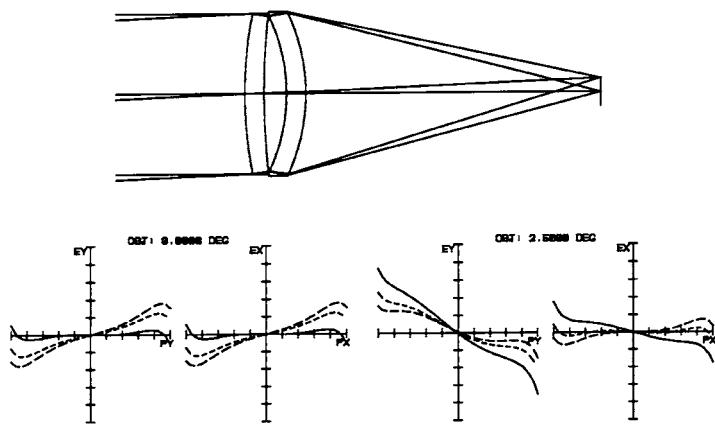


Figure 12.28
Germanium/Zinc Sulfide Lens, $f/2$ 50-mm Focal Length



Material—germanium / AMTIR 1

R1 = 64.940 cx

R2 = 114.457 cc

CT=3

Air Sp=3.5

R3 = -28.233 cc

R4 = -29.814 cx

CT = 3

BF=46.157

RMS = 0.051 λ

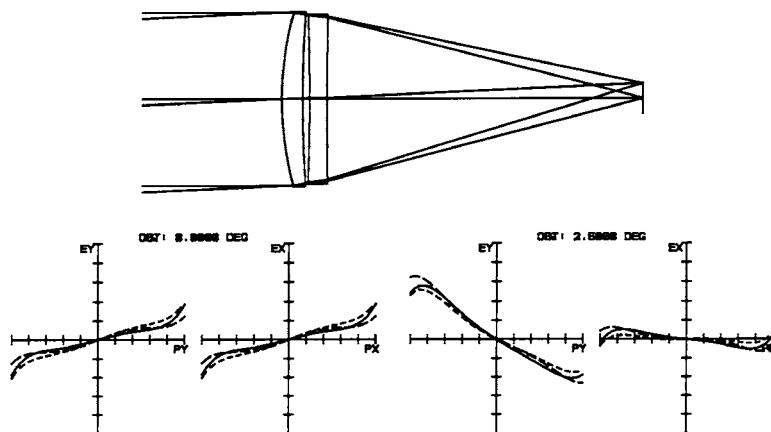
λ short = 8 μ m

λ mid = 10 μ m

λ long = 12 μ m

Figure 12.29

Germanium/AMTIR 1 Lens, $f/2$ 50-mm Focal Length



Material—AMTIR 1/ZnS

R1 = 46.876 cx

R2 = 198.763 cc

CT=3

Air Sp=1.0

R3 = -324.756 cc

R4 = 884.628 cc

CT=2.5

BF=45.566

RMS = 0.053 λ

λ short = 8 μ m

λ mid = 10 μ m

λ long = 12 μ m

Figure 12.30

AMTIR 1/Zinc Sulfide Lens, $f/2$ 50-mm Focal Length

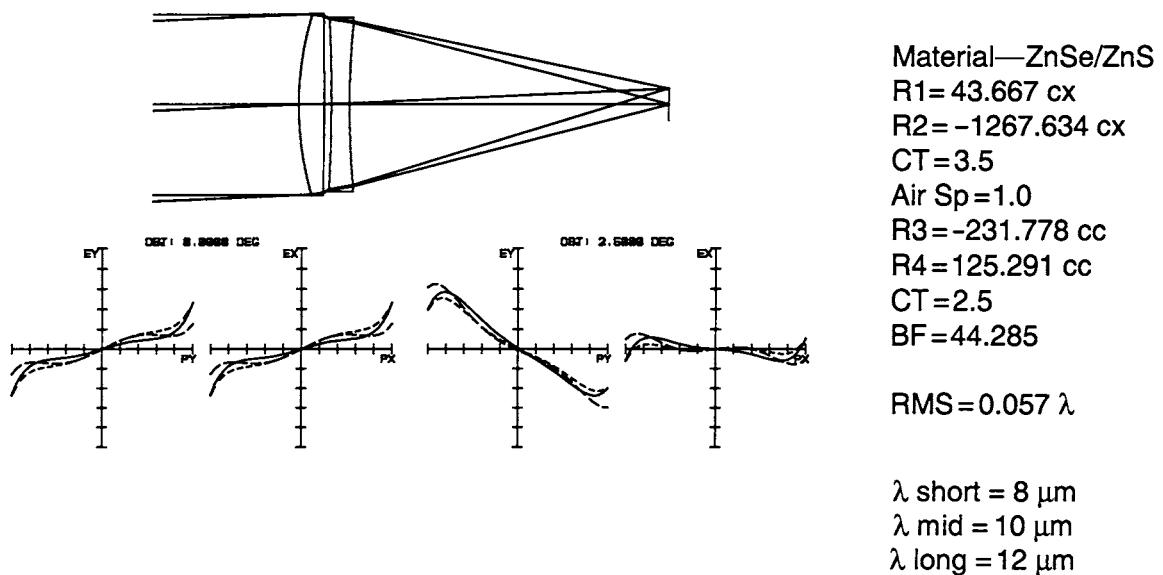


Figure 12.31
Zinc Selenide/Zinc Sulfide Lens, $f/2$ 50-mm Focal Length

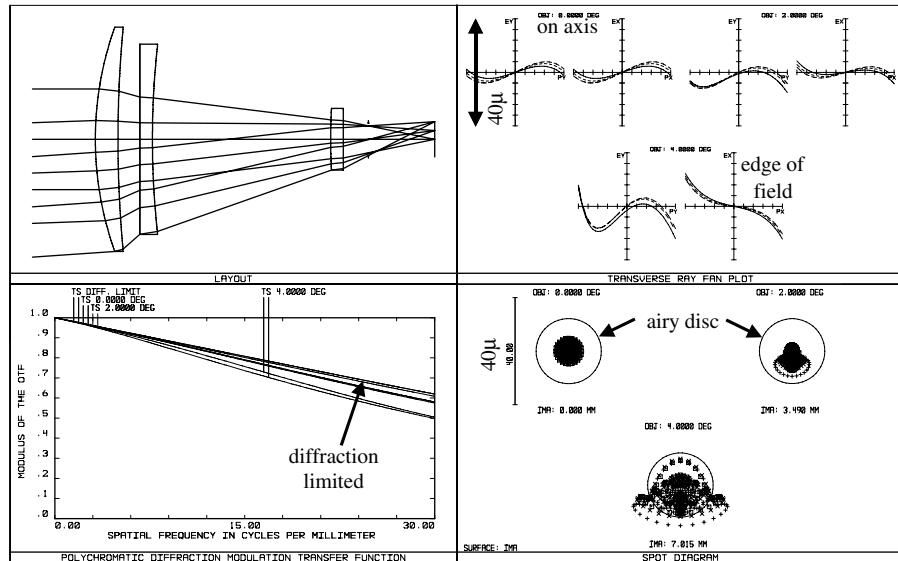
the large size of these elements with respect to the entrance pupil diameter. When you can use this nonreimaging form of design, you should do so as the design is simpler and generally performs well; however, in some situations you must use a reimaging configuration. This might be the case, for example, if you have a very tight packaging requirement.

TABLE 12.3

Relative Performance of LWIR Design Examples

Lens Construction	RMS Wavefront Error	Diffraction Limited
Germanium singlet	0.08	Nearly
AMTIR 1 singlet	0.20	No
Zinc sulfide singlet	0.85	No
Zinc selenide singlet	0.35	No
Germanium/zinc sulfide doublet	0.047	Yes
Germanium/AMTIR 1 doublet	0.051	Yes
AMTIR 1/zinc sulfide doublet	0.053	Yes
Zinc selenide/zinc sulfide	0.057	Yes

Figure 12.32
100-mm Focal Length MWIR Lens with 100% Cold Stop Efficiency



Surface	Radius	Thickness	Material
1	132.661 cx	9	silicon
2	480.403 cc	8.671	
3	Infinity	5	germanium
4	340.888 cc	70.846	
5	242.243 cx	5	silicon
6	Infinity	10	
7 stop	Infinity	26.3	
8	Image, infinity		

Optical Systems for the UV

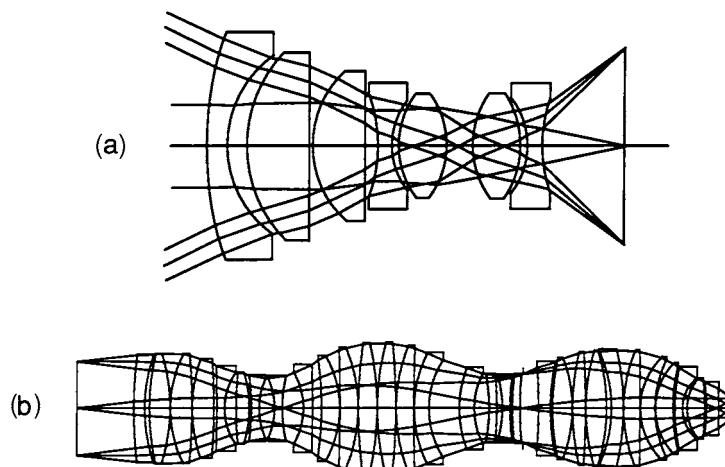
Working with systems in the UV is extremely challenging and demanding. In the thermal infrared, specifically the MWIR and the LWIR spectral bands, we found that the wavelength was 8 and 20 times the visible wavelength, respectively, which in some ways made IR lens systems more forgiving. In the UV we find that we now have a wavelength about one-half that of a visible system. In addition to the Airy disk expressed in micrometers being about one-half the f /number, we also

find a limited number of viable optical materials are available. These include fused silica, several of the fluorides (barium fluoride, calcium fluoride, and lithium fluoride), UBK7 glass, and sapphire. The index of refraction of these materials is generally not very high. Many of these materials (especially the fluorides) are very difficult to work with and have other problems such as being hygroscopic. This leads to extreme care in manufacturing and assembly, and you may need to nitrogen purge your system to prevent moisture damage. Even sodium chloride can be used in the UV, but we recommend that you “take it with a grain of salt.”

Figure 12.33 shows two deep UV lens systems, the first of which is a relatively wide-angle lens using calcium fluoride and Ultran 30 materials (the latter is no longer available). Note the relatively steep radii which is due to the inherently lower refractive indices of the materials. The second lens shown is a wafer stepper lens from a patent. This lens is similar to the Glatzel lens in Fig. 5.15, except here many more lens elements are used. Having searched the patent files extensively, this indeed is one of the more complex multielement lenses we found.

Mirror systems offer a unique advantage in the UV. Since there are no chromatic aberrations with mirrors, these aberrations are by definition zero. And since the dispersion of refractive materials is larger at lower wavelengths, using mirrors makes good sense when you can. Two very clever reflective systems are the Schwarzschild reflective microscope

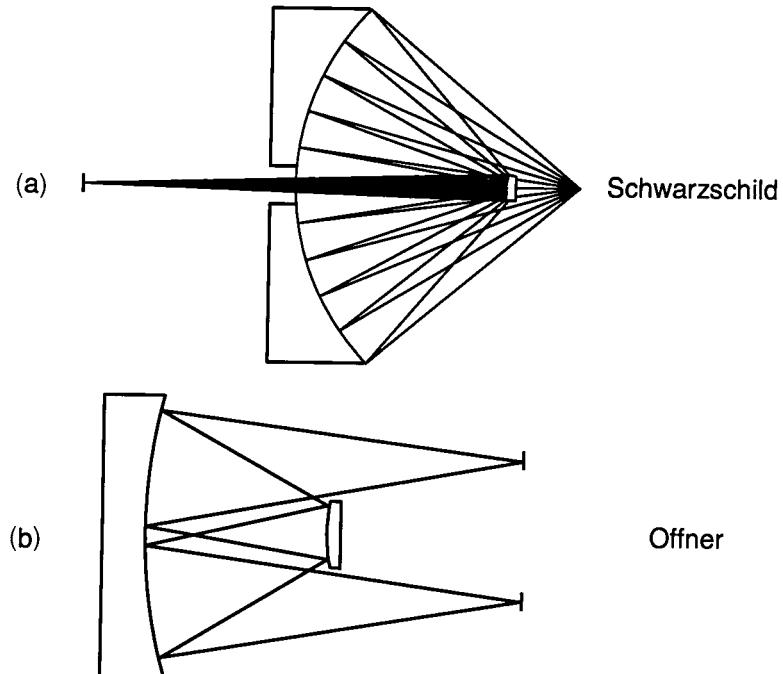
Figure 12.33
Refractive Designs for
the Deep Ultraviolet



objective and the Offner $1\times$ relay for lithography, both of which are shown in Fig. 12.34. The Offner design is especially clever in that it produces a ring field of view where virtually all orders of aberrations are corrected to zero. The original patent for this system was granted in 1973 (USP 3,748,015), and the abstract reads as follows:

A catoptric [all reflective] system for forming in accurate micro detail an image of an object at unit magnification with high resolution is provided by convex and concave spherical mirrors arranged with their centers of curvature coinciding at a single point. The mirrors are arranged to produce at least three reflections within the system and they are used in the system with their axial conjugates at said point and to provide two off axis conjugate areas at unit magnification in a plane which contains the center of curvature, the axis of the system being an axis normal to the latter plane and through said point. This combination is free from spherical aberration, coma, and distortion and, when the algebraic sum of the powers of the mirror reflecting surfaces utilized is zero, the image produced is free from third order astigmatism and field curvature.

Figure 12.34
Reflective Designs for
the Deep Ultraviolet



Note that UV systems are often prone to scattering problems. The total integrated scatter (TIS) is

$$\text{TIS} \approx \left[\frac{4 \pi \delta (\cos \theta)}{\lambda} \right]^2$$

where δ is the rms surface roughness, λ is the wavelength, and θ is the angle of incidence. This means that surface imperfections and finish, as well as bulk material scattering, can introduce unwanted stray light.

Diffractive Optics

Introduction

What Is Diffraction?

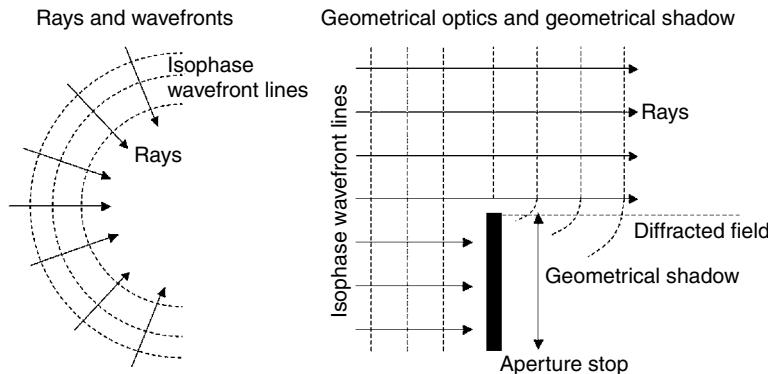
A chapter on diffractive optics within a classical optical design book is the perfect way to introduce the dual nature of light.¹ Light can be studied either through its corpuscular nature, the photon, (the basis of ray tracing and classical optical design of lenses and mirrors) or through its wave nature, an electromagnetic wave (the basis of physical optics used to model diffractive optics and other micro- or nano-optical elements, like integrated waveguides, and even photonic crystals).

In the simple knife edge example presented in Fig. 13.1, the corpuscular nature (ray tracing) accounts for geometrical optics whereas the wave nature (physical optics) nature accounts not only for the light in the optical path, but also for the light appearing inside the geometrical shadow—through diffraction at the edges of the knife edge—where no light should appear according to ray tracing.

From Refraction to Diffraction

Many diffractive elements have their counterparts in the classical realm of optical elements. However, the similarity is only superficial since their behavior under various operating configurations may be very different.

Figure 13.1
The Dual Nature of Light



As we will see later in this chapter, in many cases, diffractives are best used in conjunction with refractive or reflective optics in order to provide new and/or extended optical functionality (such as hybrid achromat or athermal singlets).

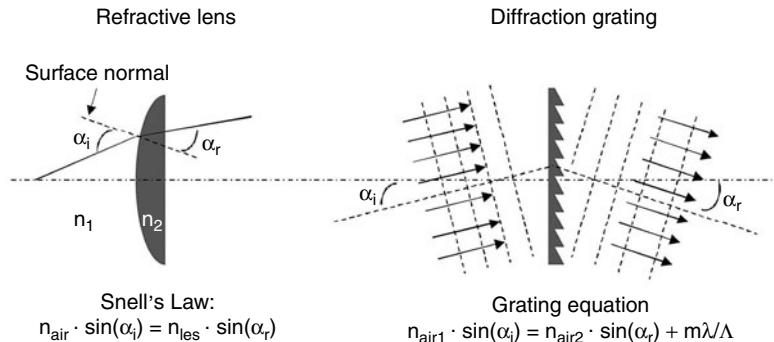
The Blazed Grating

Consider a linear blazed diffraction grating and its prism refractive counterpart, which would bend an incoming light beam into a specific—and same—direction, for a specific wavelength and a specific launch angle. The diffractive and refractive behaviors differ rapidly when one drifts away from these exact configurations. The same is true for a diffractive Fresnel lens and a refractive lens, where the dispersion occurs with opposite signs (the sign of their respective Abbe V numbers) although their focusing power and phase profiles might be exactly the same. Figure 13.2 shows both Snell's equation (ray tracing) and the grating equation (physical optics) which accounts for the amount of light bending, for a small prism and a linear blazed grating.

As one looks closer at a blazed grating, one can consider the various periods of this grating as many individual refractive microprisms, and therefore apply not only the grating equation to the entire blazed grating (array of microprisms), but also Snell's law of refraction to each individual microprism, as depicted in Fig. 13.3.

It is interesting to note that the bending angle predicted by refraction through the local microprism structures and the bending angle predicted by diffraction through the blazed grating are not necessarily the same. In effect, they are equal only in one very specific case: when the geometry of

Figure 13.2
Snell's Law of Refraction and the Grating Equation Ruling the Bending of Light



the microprism is carefully chosen (height, length, and refractive index) as shown in Fig. 13.4. Maximum diffraction efficiency is then reached for the blazed grating (which can reach 100% efficiency when both previously described effects are yielding the same angle bending).

Snell's law predicts the amount of refraction at a given optical interface, and gives thus also the expression of the refracted beam angle through the microprism (Eq. 13.1).

$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2) \Leftrightarrow \sin(\alpha + \gamma) = n_1 \sin(\alpha) \quad (13.1)$$

Physical optics, or the grating equation, gives the expression of the diffracted angle through the entire array of microprisms (Eq. 13.2).

$$\sin(\beta) = m \frac{\lambda}{\Lambda} \quad (13.2)$$

Figure 13.3
The Blazed Grating and its Microprism Array Structure

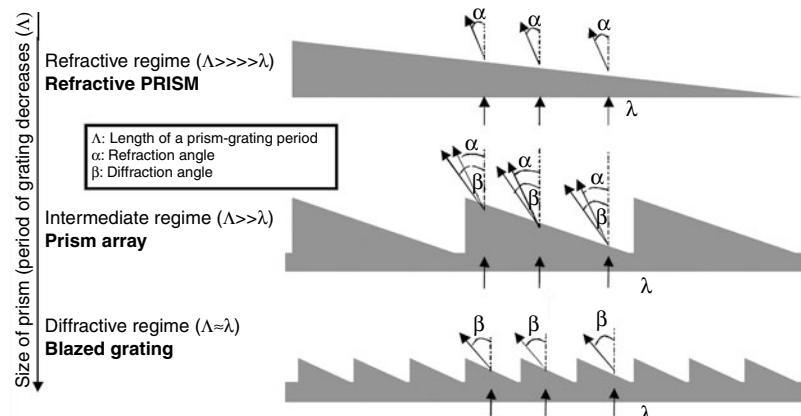
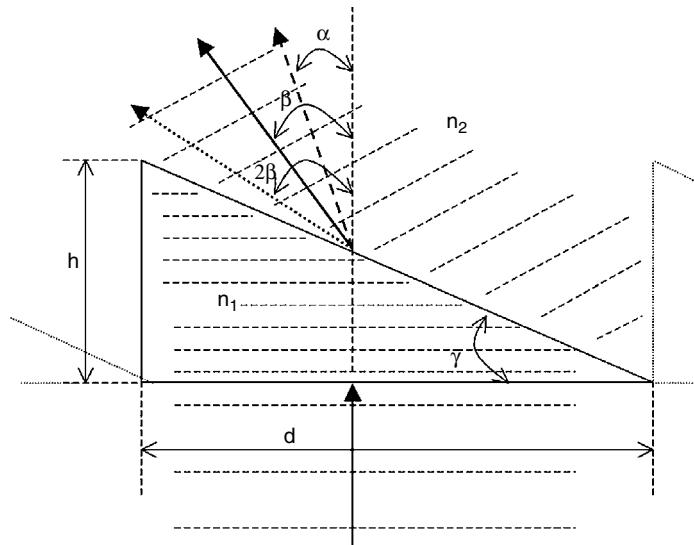


Figure 13.4

The Local Microprism Effect and the Global Grating Effect



Intuitively, the maximum efficiency will thus occur when $\alpha = \beta$ (Eq. 13.3)

$$\alpha = \beta \Rightarrow h = \frac{\lambda}{n_1 - \sqrt{1 - \frac{\lambda}{\Lambda}}} \Rightarrow h = \frac{\lambda}{n_1 - 1} \quad (13.3)$$

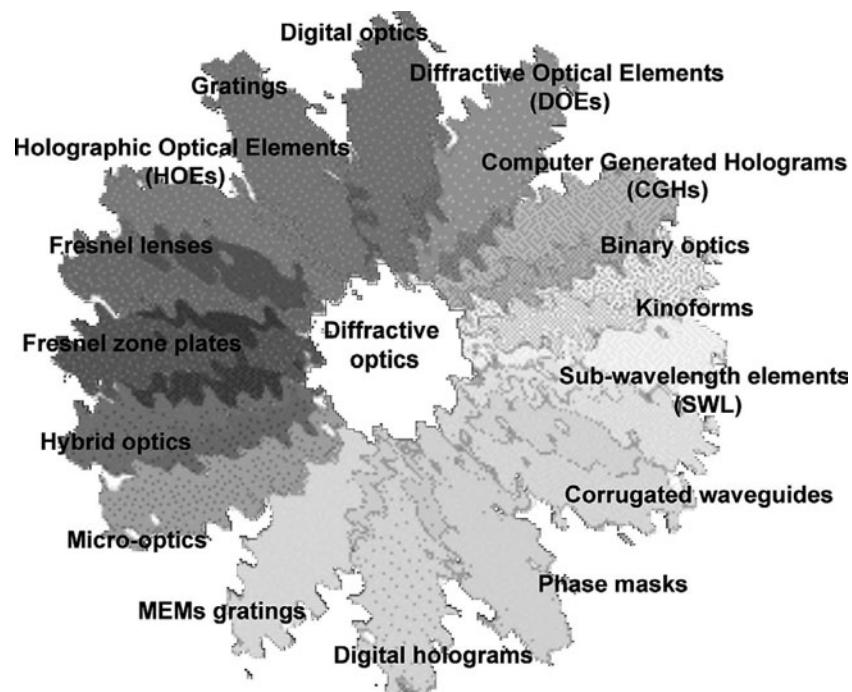
Therefore, by using the same concepts to increase the light bending efficiency, and by carefully shaping the overall grating geometry (grating period, groove height, groove angles, and refractive index), one can engineer any type of diffractive grating or diffractive lens for specific applications (aspheric lenses, circular gratings, etc. listed in the following section).

The Many Faces of Diffractive Optics

Marketing and sales managers, venture capitalists, engineers, academics as well as technical writers have given numerous names to diffractive optics in the last three decades. Some most commonly used names are binary optics or digital optics, diffractive optical elements (DOEs), computer generated holograms (CGHs), holographic optical elements (HOEs) and kinoforms. Figure 13.5 shows a compilation of those various names.

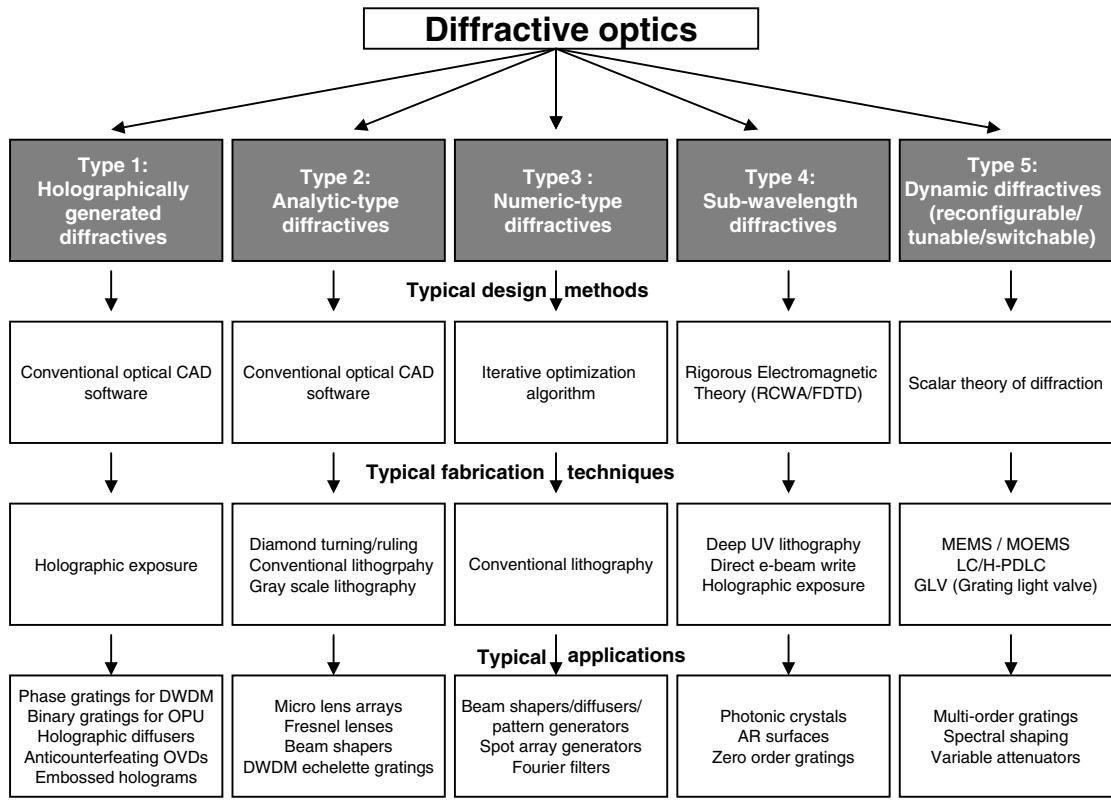
Figure 13.5

The Many Names of Diffractives



There are roughly five different groups of diffractive optics that have been reported in literature since 1967, (when Prof. Adolph Lohman first introduced the concept of “Synthetic Hologram”²), which are categorized not so much along their optical functionalities, but rather along the design techniques and the physical implementations used to produce them. Note that many different techniques can be used to design a same diffractive optical functionality. Figure 13.6 summarizes these five different types³:

1. Holographic optical elements—are referring to traditional optical holographic recording of volume phase holograms (in phase modulation materials) or surface relief holograms (in photo resist materials).
2. Analytic type diffractives—refer mostly to elements that can be designed or optimized by the means of analytical methods, like ray tracing (as it is done in most of the optical computer-animated design [CAD] tools) or by solving an analytical equation (as it is done for Fresnel lenses or gratings). These are the most common diffractives.

**Figure 13.6**

The Five Different Diffractive Elements Types

3. Numeric type diffractives—refer mostly to elements that cannot be designed or optimized by analytical methods, and require stochastic iterative optimization procedures and algorithms. These elements can implement more complex optical functions than analytic-type diffractives, but have their limitations (amount of CPU power required, need to rasterize the element in the design process, etc...). They are increasingly used in industry (see also the section “Where Are Diffractives Used?”).
4. Subwavelength diffractive elements—(SWG or subwavelength gratings) refer to elements which basic structures are smaller than the reconstruction wavelength, and are thus highly polarization sensitive and act very differently from the previous two diffractive types.

5. Dynamic diffractives—refer to all the technologies used to implement reconfigurable, tunable, or switchable diffractive optical functionalities. Note that these elements can actually incorporate any of the four previous diffractive types (as for their design and fabrication techniques). This last type of diffractive has recently gained much interest for emerging optical market and applications (especially in laser display), as we will see later in the section “Where Are Diffractives Used?”.

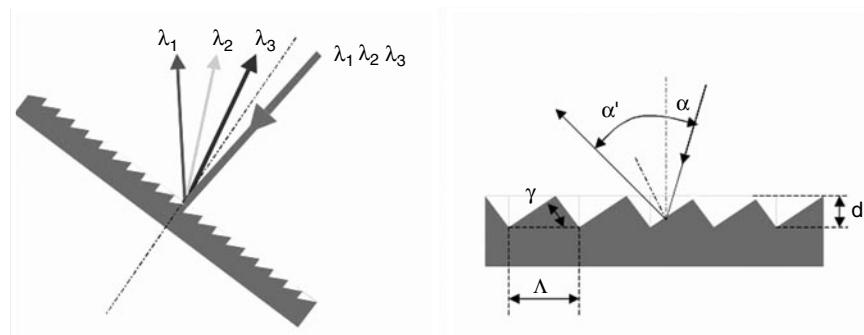
Diffraction Gratings

Linear diffraction gratings have historically been the first type of diffractive element that has been studied, fabricated and used successfully in industrial applications. They still account for most of the diffractives used in industry today (for example, in spectroscopy, dense wavelength division multiplexing [DWDM] telecom applications, optical security devices, optical data storage, optical sensors, etc...).

Linear diffraction gratings are designed with the grating equation (Eq. 13.2), and thus are analytic-type elements (Fig. 13.6). As they are rather simple in geometry, there is no need to use a special CAD tool to design such elements. They can be fabricated by a wide variety of techniques and technologies, from diamond ruling to holographic exposure and microlithography.

Figure 13.7 shows a reflective linear sawtooth grating (or échelette grating) which is used for its unique spectral dispersion characteristics (as in a DWDM telecom Demux).

Figure 13.7
A Reflective Sawtooth Linear Grating



The spectral dispersion (or strength of the wavelength demultiplexing functionality) of such a grating can be written as:

$$\frac{\partial \alpha'}{\partial \lambda} = \frac{m}{\Lambda \cos(\alpha')} \quad (13.4)$$

The resolution of such a grating can be written as:

$$R = mN_o \quad (13.5)$$

where Λ is the period of the échelette grating, λ the incident wavelength, α the incident angle, α' the diffracted angle, N_o the number of grating grooves, and m the target diffraction order considered in the application.

Note that in many cases, the optimal diffraction order is not the fundamental order, but a much higher order (diffraction orders up to 15 are commonly used in DWDM gratings).

Diffractive Optical Elements

DOEs—analytic type elements—are the most popular diffractives used today (although numeric type elements—CGHs—are getting used more and more in industry, see section “Where are Diffractives Used?” later). These elements have usually optical power (that is, lenses, unlike gratings or Fourier elements which have no optical power). The Fresnel lens is the most straightforward example of a DOE.⁴ DOEs are usually calculated in an analytical way, and in most cases one would use a classical optical CAD tool based on ray tracing to optimize an aspherical phase profile within a plane or curve (holographic or diffractive plane).

It is interesting to note that unlike its refractive lens counterpart, fabricating a highly aspherical diffractive lens bears the same price tag as fabricating a simple spherical diffractive lens.

The phase profile used to describe a diffractive element can take on many forms depending on the optical software used to design and optimize it. There are mainly three different analytical descriptions used in today’s CAD tools to describe diffractive phase profiles:

1. The traditional sag equation (Eq. 13.6):

$$\theta = \frac{2\pi}{\lambda} \left(\frac{Cr^2}{1 + \sqrt{1 - (A + 1)r^2C^2}} \right) + \sum_{i=1}^n Cr^{2i} \quad (13.6)$$

2. Rotationally symmetric elements (simple polynomial in r —Eq. 13.7):

$$\theta = \sum_{i=2}^n C_i r^i \quad (13.7)$$

3. Nonrotationally symmetric elements (general polynomial in x and y —Eq. 13.8):

$$\theta = \sum_{j=1}^m \sum_{i=1}^n C_i x^i \cdot C_j y^j \quad (13.8)$$

Once the aspheric phase profile is defined in this infinitely thin surface (which can be planar or mapped on a refractive aspherical curvature surface), the phase profile is sliced into 2π phase shift slices for the considered wavelength for maximum diffraction efficiency in the fundamental positive order, a process shown in Fig. 13.8.

An alternative and simpler way to compute the position and widths of the zones of a spherical Fresnel lens, is to calculate directly integer numbers of waves departing from the desired focal point and aiming at the DOE plane (see Eq. 13.9 and Fig. 13.9):

$$R_n = z + i \cdot \frac{\lambda}{2} \quad (13.9)$$

USEFUL PARAMETERS OF DIFFRACTIVE LENSES The numerical aperture (NA) is an important parameter of a diffractive lens and can be expressed as in Eq. 13.10 (where D is the diameter of the lens, α_{\max} the

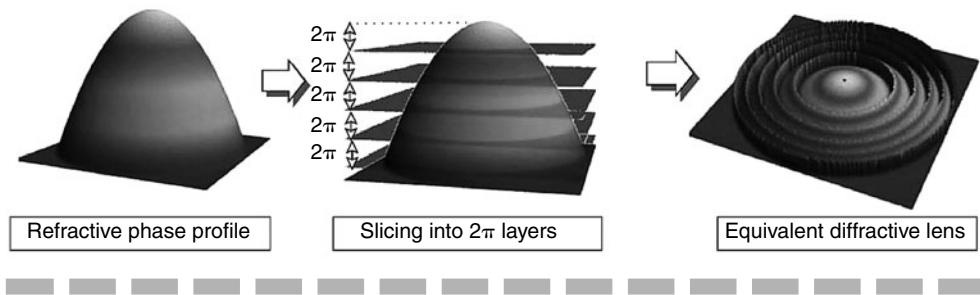
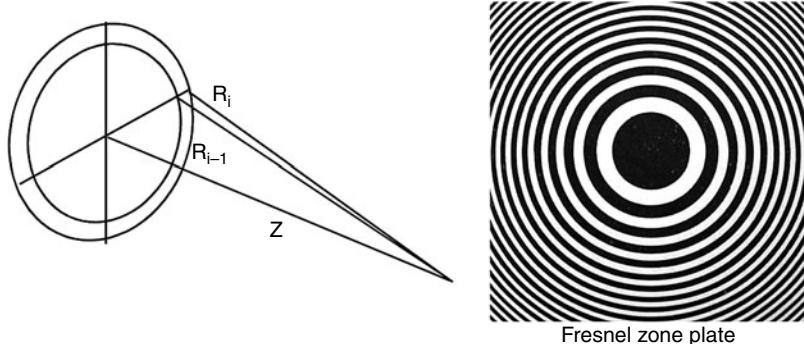


Figure 13.8

From the Initial Aspherical Phase Profile to the DOE Fringes

Figure 13.9
The Fresnel Lens
(Fresnel Zone Plate)



maximum diffraction angle at the edges of the lens, λ the reconstruction wavelength and f the focal length).

$$\text{NA} = \sin(\alpha_{\max}) = \frac{\lambda}{\Lambda_{\min}} = \sin\left(\arctan\left(\frac{D}{2f}\right)\right) \approx \frac{D}{2f} \quad (13.10)$$

The minimum fringe width δ of such a lens (or the minimal local grating period Δ_{\min} at the edges of that lens) becomes then (Eq. 13.11):

$$\delta = \Lambda_{\min} = \frac{\lambda}{\text{NA}} \quad (13.11)$$

This smallest feature is also a very important parameter, since this usually is the “go or no go” red flag when designing a lens to be fabricated by a specific technology, since *a lens is only useful if it can be actually fabricated*. For example, a 16-phase level Fresnel lens would require a $\delta/16$ minimum feature size to be fabricated. So if the minimum dimension printable on the wafer is $1.0 \mu\text{m}$, the minimum printable period of the lens will be $16 \mu\text{m}$. This parameter is also handy, as we will see later in the section on diffraction efficiency, where the efficiency of a lens is actually also a function of the local period. If the period gets too small (although printable), the efficiency can drop considerable and/or parasitic polarization effects can occur.

BROADBAND DIFFRACTIVE OPTICAL ELEMENTS Diffractive elements are very sensitive to wavelength changes (strong spectral dispersion), both in their focal plane (chromatic aberrations) and in their efficiency (see also fabrication section). Nevertheless it is, possible to optimize diffractives (for example, Fresnel lenses) to function over a wider range of wavelengths or over a set of predetermined individual

wavelengths. These lenses are commonly referred to as multiorder or harmonic diffractive lenses. The etch depth of such lenses is optimized so that its phase difference would yield an integer number of 2π phase shifts for each wavelength (see Fig. 13.10). This has the advantage of widening the fringe width, but also the inconvenient of deepening the groove depth. So precaution has to be taken when using this technique.

INTERFEROGRAM-TYPE DIFFRACTIVE OPTICAL ELEMENTS

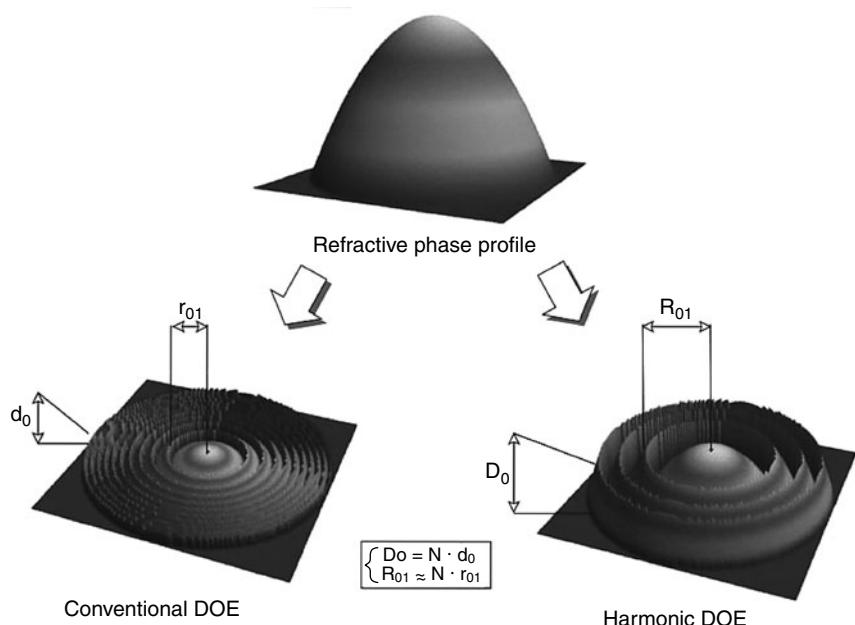
Another type of analytical element used in literature is the interferogram-type DOE⁵, in which phase function is calculated as the interference pattern of a given object wave and a reference wave (similar to holographic exposure).

As an example, consider a tilted planar reference wavefront which makes an angle α with the DOE plane, and interfering with a normal planar wavefront. The resulting interference pattern can be expressed in Eq. 13.12:

$$t(x,y) = c + 2A_0(x,y) \cdot \cos \left[\varphi_0(x,y) + 2\pi \frac{\sin(\alpha)}{\lambda} \cdot x \right] \quad (13.12)$$

Although the physical interference pattern is an intensity distribution, the resulting DOE interferogram can be recorded as a pure phase

Figure 13.10
Narrowband and
Broadband Diffrac-
tive Lens Genera-
tion from a Same
Phase Profile



element (in phase or surface relief modulation), or as an amplitude element. These elements have very remarkable properties when used in off-axis configurations as complex lenses (like toroidal, conical or helicoïdal lenses), but lack in diffraction efficiency due to their typical sinusoidal phase profile (which limit their efficiency to 33% in sinusoidal mode or 40% in binary mode—see fabrication section). In order to increase their efficiency as a surface relief element, one can fabricate the calculated fringes as blazed structures instead of sinusoidal structures.

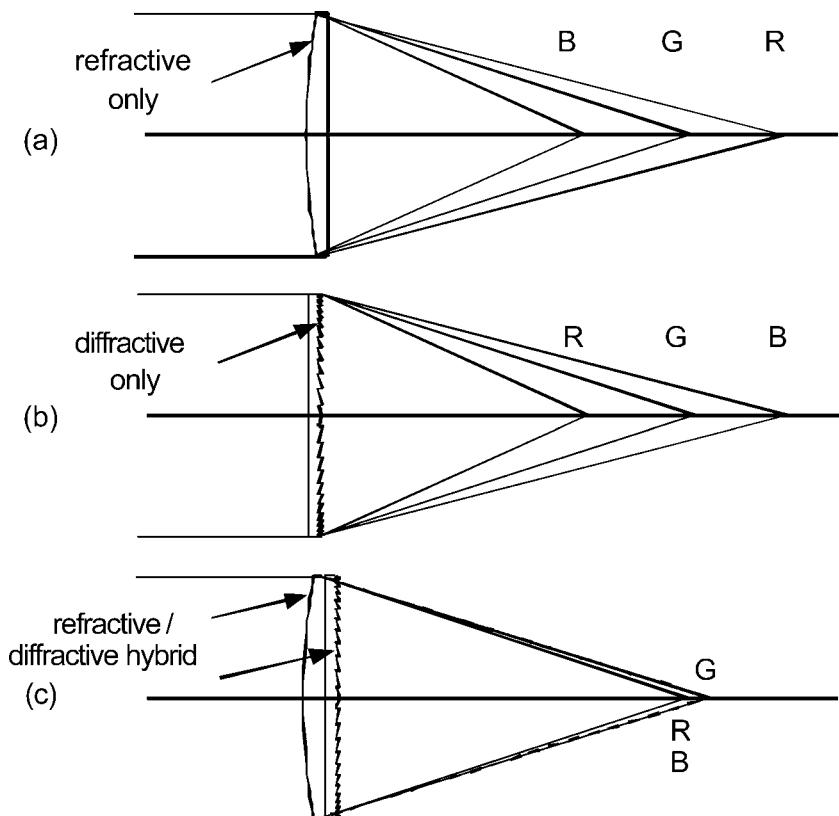
Hybrid Optical Elements

As discussed in the introduction lines of the chapter, diffractives are best used in combination with conventional optical elements, as hybrid optical elements⁶ in order to extend existing optical functionality, or introduce additional functionality. Traditional ways of achromatizing⁷ or athermalizing⁸ refractive lenses is to insert a diffractive structure carefully designed, so to balance the spectral dispersion as the Abbe V numbers of refractives and diffractives have opposite signs. As the spectral dispersion of diffractives is much stronger than for refractives, an achromatic singlet has typically a much stronger refractive power than diffractive power—about 10 times or more (see Fig. 13.11).

The section below, “What Design and Modeling Tools Should I Use?”, shows examples of parametric design of such hybrid optical elements.

Another very promising application of hybrid refractive/diffractive optics is the dual focus optical pickup (OPU) lens, which is used in most of the CD/DVD readers available in the market today. Such a lens has a convex/convex refractive, with one bearing a diffractive profile. The diffractive profile is intentionally detuned so that it produces only 50% efficiency in the fundamental positive order, leaving the rest of the light in the zeroth order. Such a lens therefore creates two wavefronts to be processed by the refractive profiles (which are always 100% efficient). When the various profiles are carefully optimized so that the zero-order combined with the refractive profiles would compensate for spherical aberrations at 780 nm wavelength through a CD disk overcoat, and the diffracted fundamental order combined with the same refractive profiles would compensate for spherical aberrations at 650 nm wavelength through a DVD overcoat, this lens is ready to pick up either CD track and DVD tracks.

Figure 13.11
Achromatization
with Hybrid Diffrac-
tive/Refractive
Singlet



Hybrid optics and hybrid optical compound lenses are also very suitable for the design of triple focus lenses to be used in the next generation BD OPUs (blue ray disks), which feature three different wavelengths (780 nm, 650 nm, and 405 nm), three different focal lengths, and three different spherical aberrations, to compensate for three different disk media overcoat thicknesses.

Computer Generated Holograms

CGHs are numeric type elements (see Fig. 13.6), which are often used as single standing optical elements, unlike analytic-type elements. The versatility of the numerical iterative optimization processes of CGHs allows the design of complex optical functionalities which could not have been implemented by classical refractive or reflective optics earlier.

CGHs are mainly used as complex optical wavefront processors, and are the tools for the development of unconventional optical applications, especially in displays, as we will show in the last section of this chapter.

There are two types of CGHs, the Fourier type and the Fresnel type.⁹ The first one projects the desired pattern in the far field whereas the second one reconstructs the pattern in the near field, and thus can be considered as a complex diffractive lens (bearing optical power as well as optical shaping/splitting).

Figure 13.12 depicts a Fourier and Fresnel CGH, which implement nearly the same optical functionality, namely splitting the incoming beam into an array of 3×3 beams. While the first one would create nine collimated beams from a single collimated beam, the second one would create nine converging beams from the same collimated beam, and therefore could be considered as a multifocus diffractive lens. Note the fringe-like pattern in the Fresnel type which does not appear in typical Fourier-type elements.

The notable difference between a standard array of 3×3 diffractive microlenses and a 3×3 multifocus diffractive lens is that the NA of the multifocus lens is much larger than for the standard microlens array (three times larger). We can therefore call such a multifocus lens a phase multiplexed microlens array, in opposition to the conventional space multiplexed microlens array.

Many different iterative optimization algorithms have been successfully developed and applied in literature to the CGH design process¹⁰⁻¹², some of them listed in the following table (Table 13.1), along with their specifications:

Figure 13.12
Fourier and Fresnel
Type CGH

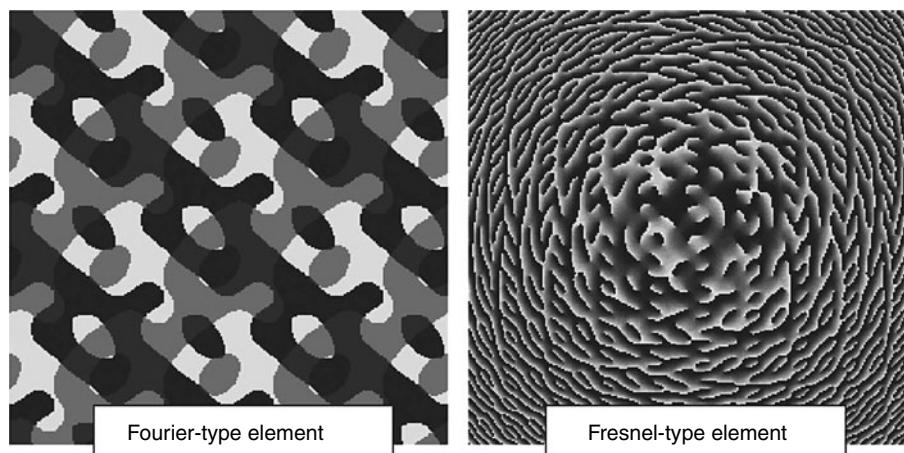


TABLE 13.1 The Various CGH Optimization Algorithms Used in Literature

Algorithm	Convergence Speed	Typical Application	Advantage	Inconvenient
Inverse FFT	Very fast	Fast display	Very simple to program	Low efficiency low uniformity, low SNR
DBS (Direct binary search)	Relatively fast	Spot array generators	Easy to program	Gets stuck into local cost function minima
Simulated annealing	Slow	Spot array generators	Very good uniformity	Slow
IFTA (GS)	Very fast	Beam shapers/displays	Very fast	Lacks uniformity
Genetic algorithms	Very slow	Spot array generators	Very versatile	Very slow, limited to small sizes
Direct analytical	Very fast	Fourier filters (band pass, edge detection,...)	Very simple to implement	Only available for Fourier filtering

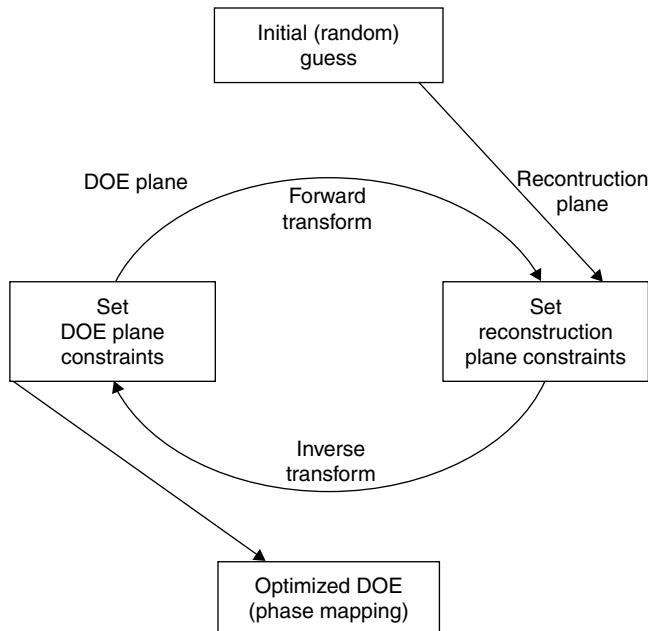
The most commonly used algorithm in industry to design both Fresnel or Fourier CGHs is an iterative Fourier transform algorithm (IFTA) called the Gerchberg-Saxton (G-S) algorithm¹³, originally developed for e-beam microscopy phase retrieval. It is both fast and very reliable (that is, it converges in most of the cases), and very easy to implement and use. However, although diffraction efficiency is optimized pretty well with the G-S algorithm, it is not the case of uniformity. Figure 13.13 shows the flow chart of the G-S algorithm.

Once the complex data of the CGH have been calculated (optimized), the optical design engineer has to choose a specific encoding technique in order to encode the phase, amplitude or complex information into a substrate.¹⁴

We listed below some of the most commonly used CGH optical functionalities¹⁵:

- Optical image processing and Fourier filtering (edge detection, low- and high-pass filters, etc.)
- Optical interconnections via spot array generators
- Far field pattern generation for two-dimensional display or focus control

Figure 13.13
The G-S Iterative CGH Optimization Algorithm (IFTA Type)



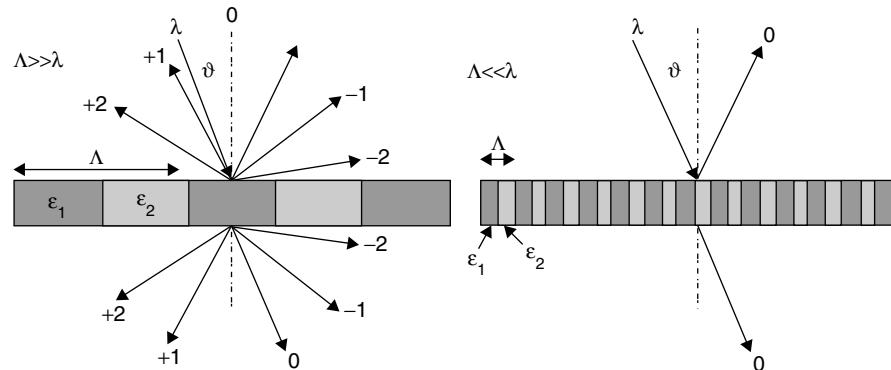
- Beam shaping (for example, Gaussian to top hat)
- Beam splitting
- Beam sampling
- Structured illumination generation for machine vision

Subwavelength Diffractives and Photonic Crystals

Subwavelength (SW) diffractives (type 4 elements in Fig. 13.6 classification) are microstructured diffractive elements in which the smallest features (or local grating periods) are smaller than the reconstruction wavelength ($\Lambda/\lambda \ll 1$), such as only the zero—backward and forward—orders are propagating, all other higher diffraction orders are evanescent.

As the incident wave cannot resolve the SW structures, it sees only the spatial average of its material properties (effective refractive index). Figure 13.14 shows the differences in diffracted and propagating orders between a standard multiorder diffractive and a SW diffractive.¹⁶

Figure 13.14
Multiorder Diffractives
and SW Diffractives
Operations



A zero-order grating—or SWG—(where no other diffracted order is propagating) is defined when Eq. 13.13 holds true:

$$\frac{\Lambda}{I} < \frac{1}{\text{Max}\{\sqrt{\varepsilon_1}, \sqrt{\varepsilon_{III}}\} + \sqrt{\varepsilon_1 \sin \theta_{\max}}} \quad (13.13)$$

Note that SW diffractives can be implemented either as thin surface relief phase elements or as index modulated holographic elements. These elements also show very strong polarization dependence, unlike standard diffractive with much larger feature sizes.

Applications requiring SW diffractives include mainly:

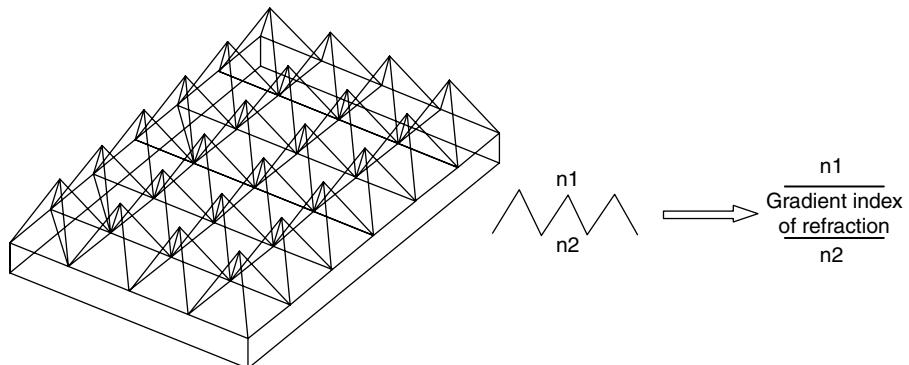
- Polarization sensitive elements (polarization splitting, polarization combining)
- Antireflection (AR) surfaces
- High-resolution resonant filters (in reflection or transmission)
- Integrated waveguide gratings (with Bragg reflection or coupling effects)
- Phase plates (in reflection and transmission)

Figure 13.15 shows such an antireflection surface composed by a two-dimensional array of pyramidal structures with SW periods.

Another application where SW diffractives show their full potential is the encoding of smooth phase profiles as small binary SW structures.¹⁷ For example, the design and fabrication of SW blazed Fresnel lenses or sawtooth gratings have been reported. The incident wave does not

Figure 13.15

Example of a Two-Dimensional SW Period Grating Producing an Antireflection Surface



resolve anymore the local binary phase modulation, since its wavelength is much larger than these structures, but sees the overall effective medium picture. Thus, the incident wave considers the material as a smoothly varying phase element. The effective medium theory (EMT) is applied to such elements in order to model them.

Holographic Optical Elements

HOEs—type 1 elements in the Fig. 13.6 classification—are considered as digital diffractive elements in the sense that the recording setup is usually designed by a CAD tool, although the fabrication remains mostly analog (holographic recording). Figure 13.16 shows the holographic recording set-up of a simple linear HOE grating.

A holographic grating is considered as thick or thin in Bragg incidence when its quality factor Q is respectively larger than 10 or lower than unity. For values of Q in between 1 and 10, the grating behaves in an intermediate state. Q is defined in Eq. 13.14 as:

$$Q = \frac{2\pi\lambda d}{n\Delta^2 \cos \alpha} \quad (13.14)$$

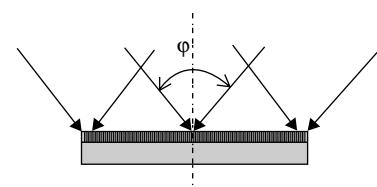
Figure 13.16

The Recording of a Linear Grating HOE as a Surface Relief Element in Resist

Holographic grating example:

The period Λ of the interference fringes generated when two beams of wavelength λ intersect at an angle of φ is given by:

$$\Lambda = \frac{\lambda}{2 \sin \frac{\varphi}{2}}$$



where d is the thickness of the grating, Λ its period, n its average refractive index and α the incident angle.

See also the section “How Are Diffractives Fabricated?”, later in this chapter for further information about holographic recording of HOEs.

What Design and Modeling Tools Should I Use?

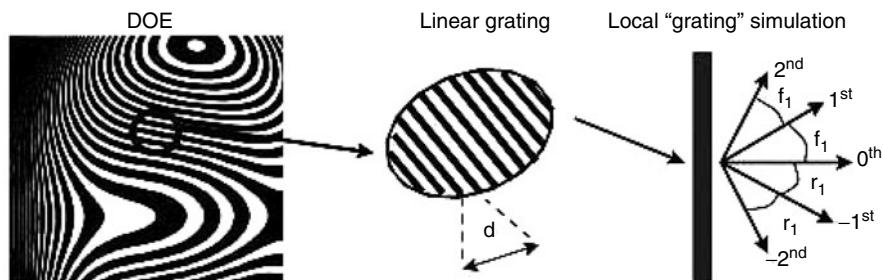
In the previous section we have described the various design techniques used today to calculate and optimize analytic as well as numeric type diffractives. In this section, we will focus on the various tools available to the optical engineer to accurately model the behavior of these diffractives (CAD tools used for modeling effects of illumination and/or opto-mechanical tolerancing, effects of various fabrication techniques and technologies and related systematic fabrication errors—see also next section on fabrication).

Ray-Tracing-Based Tools

Since diffractives are very often used in conjunction with other optical elements (refractives, reflectives, graded index, etc.), these simulation tools have to be able to interface with standard optical modeling tools, for the most based on ray tracing. Therefore, many optical CAD tools available on the market use simple ray tracing algorithms through diffractives as is the diffractive it would be special refractive element. The ray tracing uses the local grating approximation (LGA) for any diffractive element constituted by smooth fringes. Note that this is possible only if the element is constituted of fringes like a DOE or Fresnel CGH, a Fourier CGH could not be modeled this way. Figure 13.17 shows how the local grating approximation uses the grating equation in order to predict the direction of the ray passing through a particular area of the DOE.

The local angle of diffraction is given by the grating equation at the equivalent local grating location, and the efficiency at that location is computed as a function of the diffraction order considered, the groove depth, the wavelength, and the number of phase levels.

Figure 13.17
Ray Tracing by LGA
of an Aspherical Diffractive Element



Note that this technique does not inform about crucial aspects like real diffraction efficiency, multiorder diffraction, and resulting multi-order interferences. Therefore, this technique can only be used effectively for 100% efficient elements, like blazed Fresnel lenses or gratings.

This technique is thus best suited for the modeling of hybrid refractive/diffractive elements, where the diffractive element is usually fabricated by diamond turning, and thus yields very smooth fringes and high diffraction efficiency. This method constitutes the vast majority of diffractive optics modeling tools available in optical CADs on the market today.

Other similar techniques compare the diffractive to a refractive element which would have an infinite refractive index, and thus being infinitely thin, the Sweatt model.¹⁸

Scalar-Diffraction-Based Tools

In order to simulate the effects of multiorder diffraction in the scalar domain⁹, as well as the effects of fabrication techniques and systematic fabrication errors, it is best to use a physical optics approach, and implement scalar-diffraction-based propagators, which consider in parallel all propagating orders through the diffractive. Such scalar diffraction propagators can model any type of diffractive structures, composed of fringes or not (DOEs and CGHs), as analog, binary or multilevel surface relief elements. This is true only in the realm of validation of scalar diffraction theory, that is, as long one is in the paraxial regime (low NAs [see Eq. 13.10], low angles, and smallest structures much larger than the wavelength, without any polarization effects).

Helmholtz's wave equation, with Huyghen's principle of secondary sources over the wavefront's envelope, injected in Green's function is the major foundation of scalar theory, and gives rise to the Helmholtz and

Kirchhoff integral theorem. The Rayleigh-Sommerfeld diffraction formulation for monochromatic waves follows, and gives rise to the Fresnel and then Fourier approximations of the diffraction through a thin planar screen in the far and near fields.

These two formulations are the basis of most of the physical optics modeling tools used today. Equation 13.15 summarize these two formulations:

$$\left. \begin{array}{l} \text{Fourier approximation (far field)} \\ \text{Fresnel approximation (near field)} \end{array} \right\} \begin{aligned} U_D(u, v) &= \iint_{-\infty}^{\infty} U(x_1, y_1) \cdot e^{-j2\pi(ux+vy)} \cdot dx \cdot dy \\ U_D(x_0, y_0) &= \left(\frac{e^{jkf}}{j\lambda f} \right) \iint_{-\infty}^{\infty} U(x_1, y_1) \cdot e^{j\frac{\pi}{\lambda f}((x_0-x_1)^2 + (y_0-y_1)^2)} dx_1 \cdot dy_1 \end{aligned} \quad (13.15)$$

Note that the Fresnel approximation can be described in two different ways, as a direct integral here or as a convolution. When using complex two-dimensional fast Fourier transform (FFT) algorithm, one can rewrite the previous equations and implement them in a numerical tool.

As an application example, we consider here a hybrid refractive/diffractive lens used in an infrared (IR) digital camera. The lens has a spherical/aspherical convex/convex surface with an aspherical diffractive surface on the aspherical refractive surface (second surface). The lens is optimized and modeled in a standard optical CAD tool from the market as far as ray tracing is concerned.¹⁹

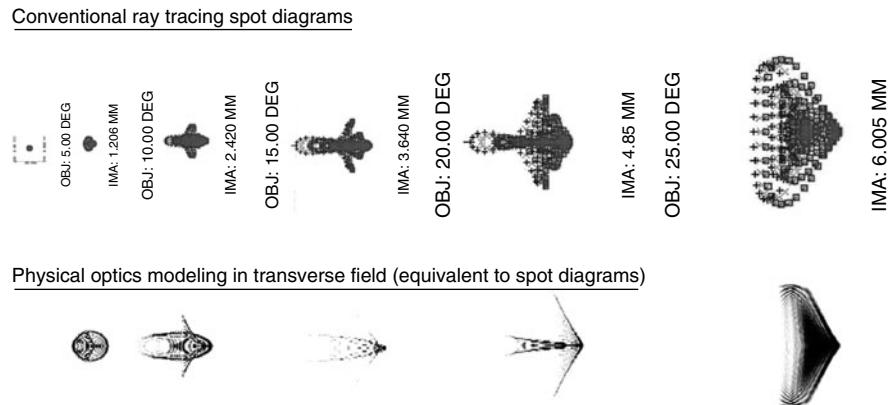
We will perform a classical lens field of view analysis, which means that we will launch a collimated wavefront on this lens under increasing angles and compute the resulting focus plane, both with standard ray tracing tools and with a numerical Fresnel propagator based on the previously described analytical formulation of diffraction in the near field.

The resulting lateral spot diagrams for incident angles from 0° to 25° are shown in Fig. 13.18.

Both modeling tools (ray trace and scalar numerical Fresnel propagators) agree well on all spot diagrams, even with angles up to 25°. However, the resolution in the numerical analysis tool when using the scalar model is much greater than when using only geometrical ray trace. Also, effects of multiple diffraction orders interferences (see the fringes created in the reconstructions) can be observed, which cannot be demonstrated with conventional ray trace algorithms.

Figure 13.18

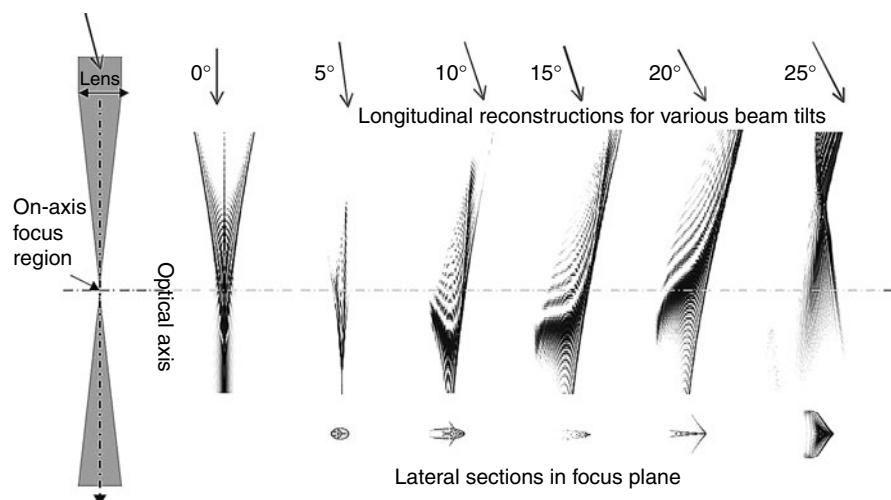
Ray Traced Spot Diagrams and Corresponding Physical Optics Modeling of a Hybrid Refractive/Diffractive Lens—a Comparison in the Lateral Plane



Scalar propagation of diffracted/refracted/reflected wavefronts gives actually much more flexibility in modeling diffractives (and any other microoptical elements), in the sense that not only lateral spot diagrams can be computed, but also longitudinal intensity profiles, as shown in Fig. 13.19. Actually, the reconstruction can be computed on any surface, planar or curved, even in a three-dimensional volume. This gives a much deeper insight on the modeling aspects, which was not possible with only ray tracing.

Figure 13.19

Longitudinal Reconstructions via Numerical DFT-Based Scalar Propagators



FFT- VERSUS DFT-BASED NUMERICAL PROPAGATORS FFT based propagators, however have yield some severe drawbacks, which are namely limitations on the size of the sampled field, location of the reconstruction plane as well as amount of off-axis of that reconstruction plane. Discrete Fourier transforms (DFT) are thus used in order to implement the Fresnel and Fourier approximations of the Rayleigh-Sommerfeld integral (Eq. 13.15). Actually, the exact Rayleigh-Sommerfeld integral can also be implemented by using DFT algorithms, which is not the case with FFT algorithms.

The main disadvantage of DFT based propagators is the CPU required to compute the numerical reconstruction. There is actually an exponential factor between the speed of FFT based and DFT based numerical propagators. The numerical reconstructions in Fig. 13.19 have been computed by using a DFT based propagators.

Rigorous Electro-Magnetic Modeling Techniques

Scalar theory of diffraction as expressed previously as a thin phase approximation is an accurate tool for elements which have smallest feature dimensions on the order of the reconstruction wavelength.^{20,21}

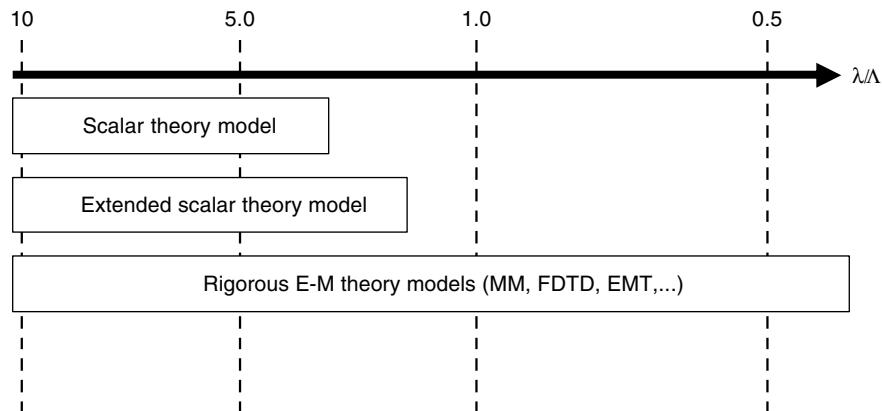
In order to account for the coupling effects along the boundary of the microstructures, rigorous electromagnetic modeling techniques should be used. Many different techniques have been reported in literature. The modal method (MM), the finite difference time domain (FDTD), and the effective medium (EM) theories are amongst them.

In between the realm of validity of scalar theory of diffraction and the more rigorous EM theories, extended scalar theories have also been reported, which deal mainly with the shadowing effect for high angles in surface relief elements. Figure 13.20 shows the three modeling regions and which technique should be applied.

When contemplating the possibility to use a rigorous modeling technique to design and model a diffractive element, it is important for the optical engineer to remember that due to high CPU time and complex implementation of the numerical tools, anything more complex than a linear grating is nearly impossible to model. Therefore, rigorous theories are not often used to design DOEs or CGHs.

It is also interesting to remember that the diffraction angles (and thus the position and geometry of the optical reconstruction) predicted by

Figure 13.20
Realm of Validity for
Scalar and Rigorous
Diffraction Theories



scalar theory still hold true even though if outside of the realm of scalar theory. Only the prediction of the diffraction efficiency is different from those predicted by the more accurate rigorous theories.

Parametric Design Example of Hybrid Optics via Ray Trace Techniques

In order to illustrate hybrid diffractive/refractive optics design through ray trace techniques with a conventional optical CAD tool and demonstrate the relative merits of these designs, we will show several representative examples. The specifications for the design example are as follows:

Entrance pupil: diameter 25 mm

Field of view on axis only

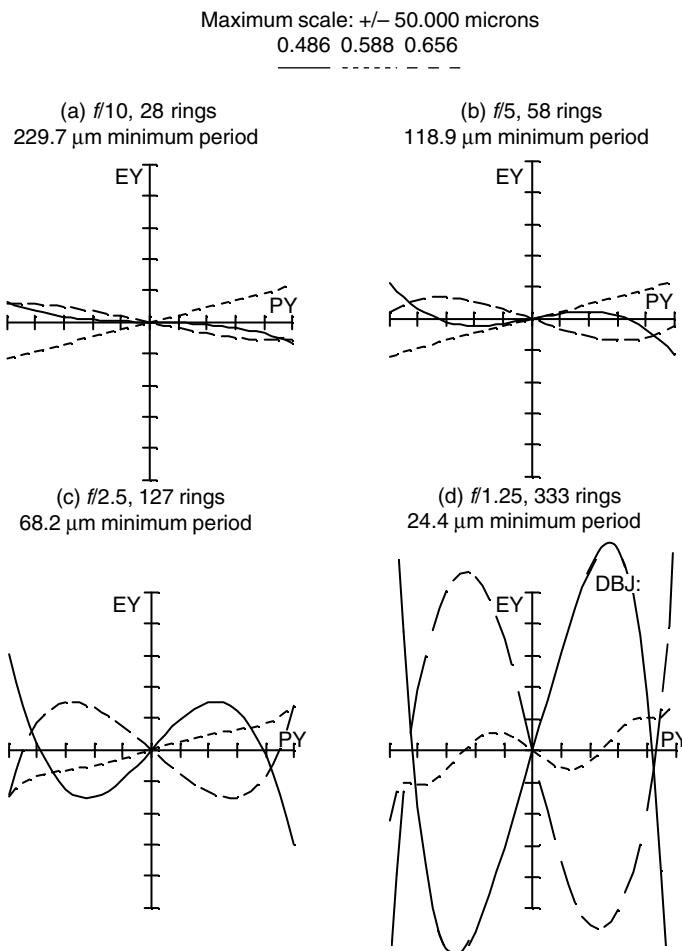
Wavelengths C (656.3 nm), D (587.6 nm), F (486.1 nm)

f#: f/10, f/5, f/2.5, f/1.25

Figure 13.21 shows the transverse ray aberration curves for a *f*/10 hybrid singlet per the specifications above. The substrate material is BK7 glass, and very similar results would arise for acrylic, which would be a fine material choice if the element were to be mass produced.

The diffractive surface is located on the second surface of the lens. The spacing or separation between adjacent fringes was allowed to vary with respect to the square as well as the fourth power of the aperture radius, or y^2 and y^4 , where y is the vertical distance from the vertex of

Figure 13.21
Hybrid Refractive/Diffractive Achromatic Singlet as a Function of $f\#$

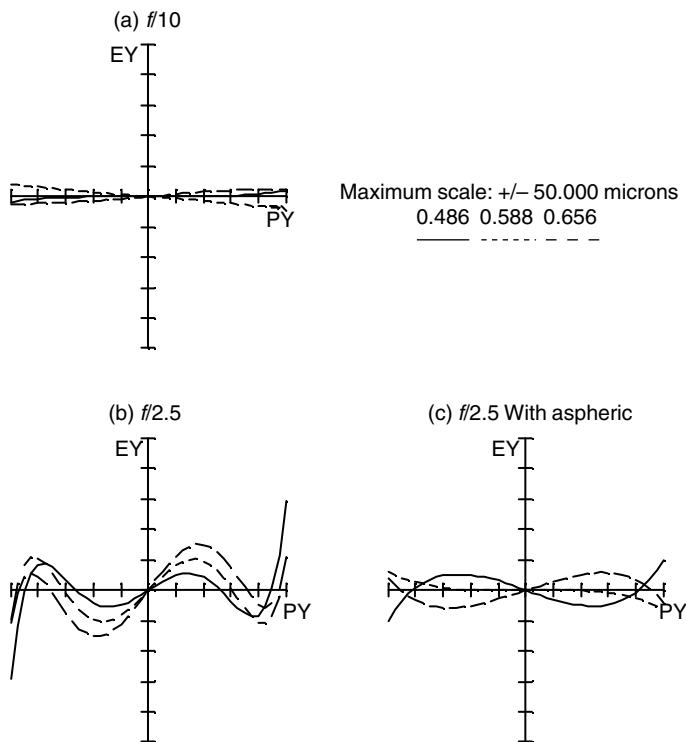


the surface perpendicular to the optical axis. The quadratic term allows for the correction of the primary axial color whereby the F and C light (blue and red) are brought to a common focus. The fourth order term allows correction of the third order spherical aberration as well. The resulting ray trace curves show the classical performance typical of an achromatic doublet. Since the lens is of relatively high $f\#$, the spherical aberration at the center wavelength is fully corrected. There is a residual of spherochromatism which is the variation of spherical aberration with wavelength. This residual aberration is due to the fact that the dispersion of the diffractive is linear with wavelength whereas the material dispersion

of BK7 glass is nonlinear. The resulting surface has 28 rings with a minimum period of 229.7 μm . The insert for injection molding this surface could be easily diamond turned. Figure 13.22 shows similar results for hybrid singlets that are $f/5$, $f/2.5$, and $f/1.25$, respectively. Note that as the $f\#$ gets lower and lower the higher order spherical aberration increases, and the spherochromatism increases as well, to a point where the spherochromatism is the predominant aberration. The number of rings and the minimum fringe period in the diffractive are listed.

Note that these data do not scale directly, as we can do with conventional optical designs. As a diffractive optical element is scaled down in focal length while maintaining its $f\#$, a true linear scaling of all parameters (except of course the refractive index which is unitless, and thus does not scale), is not correct. This is because we need to maintain the fringe depths to create a total of 2π phase shifts for the target wavelength, and a linear scaling would result in only 1π phase shift. Thus, we find that for a 0.5 scaling, we end up with one half of the number of fringes with

Figure 13.22
 $f/10$ and $f/2.5$ Classical Achromatic Doublets



approximately the same minimum fringe period. For example, if we were interested in a 125 mm diameter $f/2.5$ hybrid, we would scale the radii, thickness, and diameter by $0.5\times$ from the 25 mm starting design. However, the number of fringes will decrease by a factor of 2, with essentially the same minimum fringe period. It is highly recommended to reoptimize any lens or lens system containing one or more diffractive surfaces after scaling in order to assure that the surface prescription is correct.

It is feasible to manufacture diffractive surfaces as well as binary surfaces with minimum periods of several microns or less, however it is best to discuss specific requirements with the foundry prior to finalizing the design (see also the section "How Are Diffractives Fabricated?"). Figure 13.22 shows, for comparison, the performance of classical $f/10$ and $f/2.5$ achromatic doublets using BK7 glass.

The results are somewhat improved over the hybrid solution. Note that at the lower $f\#$ of $f/2.5$ (b) the spherical aberration of the achromatic doublet is becoming a problem, and we show an aspherical design in (c).

Figure 13.23 shows graphically the relationship between the lens $f\#$, the number of fringes, and the minimum fringe period for the hybrid designs in Figure 13.21. As we would expect, the lower the $f\#$, the larger the number of fringes and the smaller the minimum period. Figure 13.24 shows several design scenarios all for a constant $f/5$ single element lens.

Figure 13.23

Number of Fringes
and Minimum Period
as a Function of $f\#$
for a Hybrid Singlet

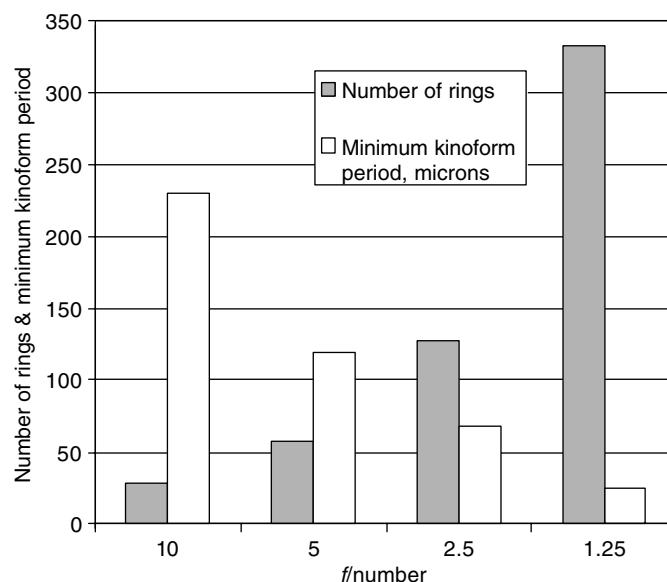
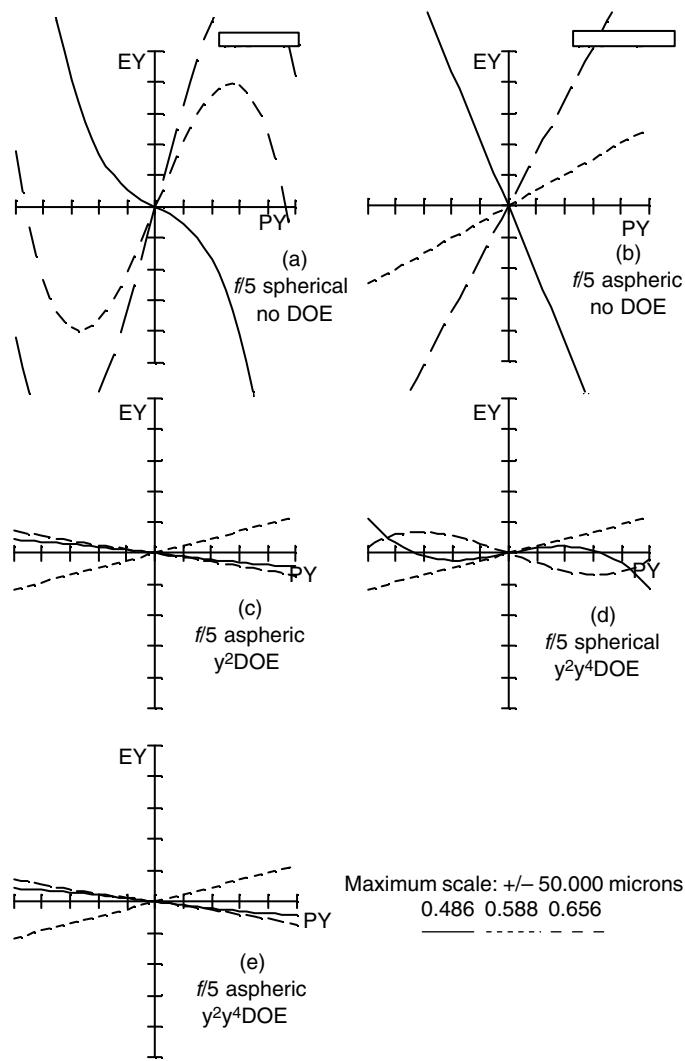


Figure 13.24
Performance of f/5
Hybrid Singlets with
Different Surface
Descriptions



For reference, (a) and (b) both have no diffractive surface, however (b) does have an aspheric surface for correction of spherical aberration. The primary axial color is the same in both lenses, and quite large as expected. Figure 13.24c has an aspheric surface for spherical aberration correction and a quadratic diffractive surface for correction of the primary axial color. Figure 13.24d is all spherical with a quadratic and a fourth order diffractive fringe width variation. It is interesting that this solution is very similar to (c), except that this solution has more spherochromatism than

the solution with the aspheric surface. Finally, (e) allows both an asphere as well as a quadratic and fourth order diffractive fringe width variation. The aspheric along with the quadratic diffractive fringe width variation of (c) was so well corrected that no further improvement is possible here.

One of the more interesting observations is that the aspheric surface along with the diffractive surface allows for the correction of both the spherical aberration as well as the spherochromatism. The all-diffractive surface with the quadratic and the fourth order fringe width variation has a residual of spherochromatism. The reason for this subtle difference is that in the aspheric case the spherical aberration correction and the chromatic aberration correction are totally separate from one another thereby allowing better performance. In the all-diffractive design we are more constrained and do not have sufficient variables to eliminate the spherochromatism as well as the spherical aberration and the primary axial color.

How Are Diffractives Fabricated?

Depending on the target diffraction efficiency and smallest feature size in the diffractive element, the optical engineer has the choice of a wide variety of fabrication technologies ranging from simple grating ruling to complex gray scale optical lithography.

The flowchart in Fig. 13.25 summarizes the various fabrication technologies as they have appeared chronologically.²²

For a single optical functionality, for example, a spherical lens, the optical engineer can decide to use many different fabrication technologies which have their specific advantages and limitations, mainly in terms of diffraction efficiency. Figure 13.26 shows six different physical implementations of the same diffractive lens, with their respective fabrication technology and their respective diffraction efficiencies (both theoretical and practical).

This invites us directly to the definition of diffraction efficiency, one of the most critical criteria when it comes to fabricate a diffractive element.

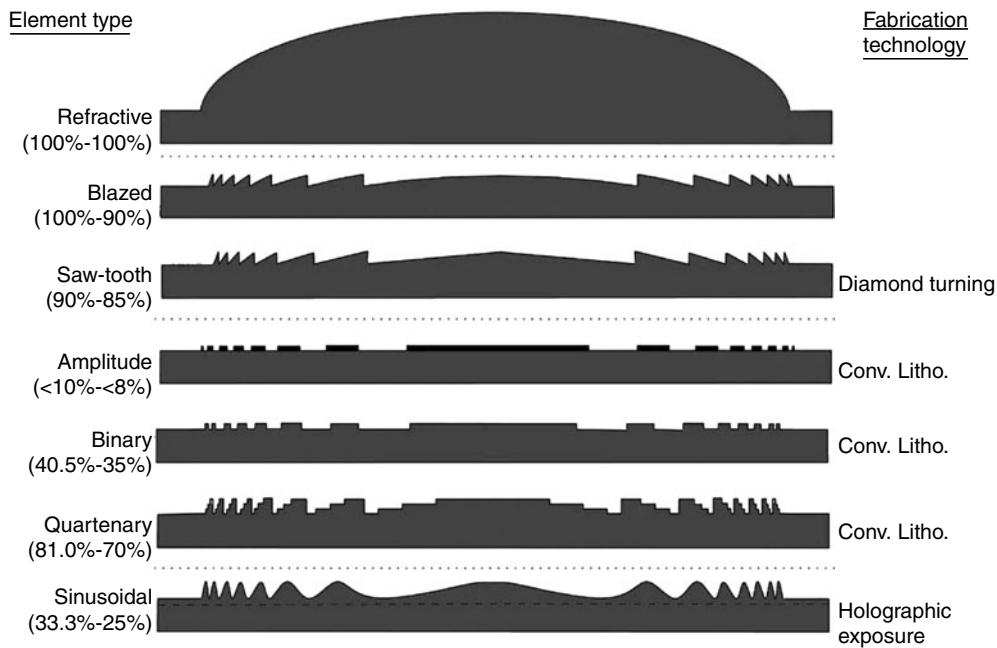
Holographic Exposure

As discussed in the previous section, holographic exposure of HOEs can yield a wide variety of optical functionalities, either as a thin surface relief HOEs (by the use of photo resist spun over a substrate) or as a

Figure 13.25

Chronological Fabrication Techniques and Technologies for Diffractives

Historical overview		
Lithography boom		
1785	Rittenhouse	- Grating made of hairs
1882	Rowland	- Ruling engines
1910	Wood	
1948	Gabor	- Holography
1962	Leith/Upatnieks	- Holography Optical Elements (HOEs)
1969	Lohmann/Brown	- Computer Generated Holography (photoreduction)
1972	D'Auria/Huignard	- Photo-refractive materials
1975	-	- CNC precision diamond ruling/turning
1983	Gale/Knop	- Direct analog laser write in resist
1985	Arnold	- Direct binary e-beam write
1989	Swanson/Weldkamp	- Multilevel optical lithography
1992	Lee/Daschner	- Direct analog e-beam write
1993	Potomac	- Direct analog Excimer laser ablation
1996	Canyon materials	- Gray scale masking lithography
2000	Digilens, SBG Labs	- Dynamic holography
2001	DOC, DS,	- Wafer scale hybrid integration
2003	-	- Dynamic digital diffractives (MEMS and others)

**Figure 13.26**

Six Different Physical Encoding Schemes for a Same Fresnel Lens

more complex Bragg grating structures in a volume hologram emulsion. Figure 13.27 shows an example of holographic recording of a simple transmission HOE lens by using two source points at a specific wavelength required by the material and playing back the hologram at another wavelength, dictated by the application.

More complex optical functionalities can be recorded by using a master diffractive element like a DOE or a binary or multilevel CGH. In this double recording process, a single diffraction order is used from the CGH (which diffracts multiple propagating orders) to generate the object beam, all other diffraction orders present are blanked out. Figure 13.28 shows such a CGH/HOE recording process. Although the CGH generates many orders, the resulting HOE will generate only a single diffraction order without altering the initial optical functionality (if the recording has been done properly in the Bragg regime angles).

Table 13.2 shows the various holographic materials used in industry and their specifications.

Although a very versatile and efficient method, holographic recording of volume HOEs cannot be used in many cases, where materials with long lifetime are necessary (due to temperature swings, vibrations, humidity, or daylight/UV exposure), or simply when mass replication of cheap diffractive elements are necessary.

Example:
Off-axis transmission lens

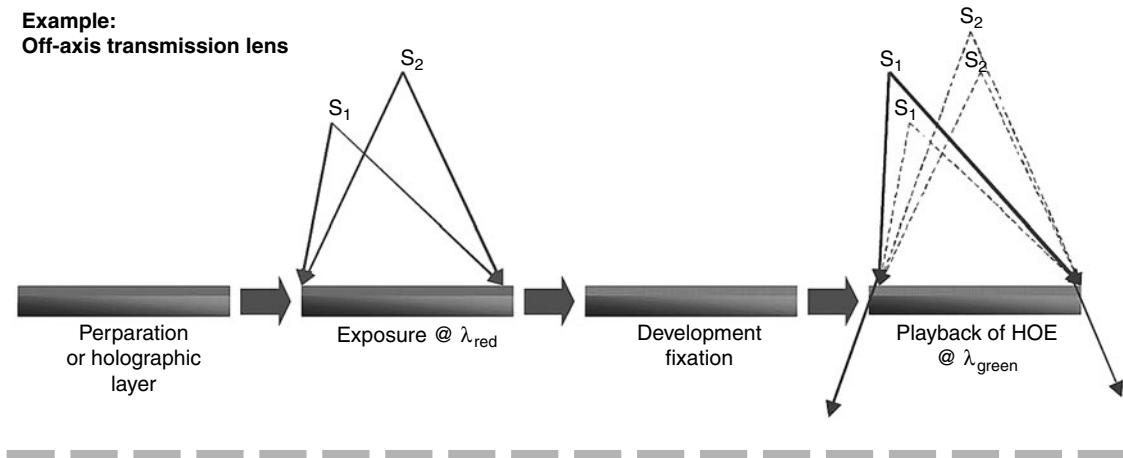


Figure 13.27

Standard Holographic Recording of an HOE Lens in a Holographic Emulsion

Figure 13.28
Surface Relief CGH
Recording Process
into a Volume HOE

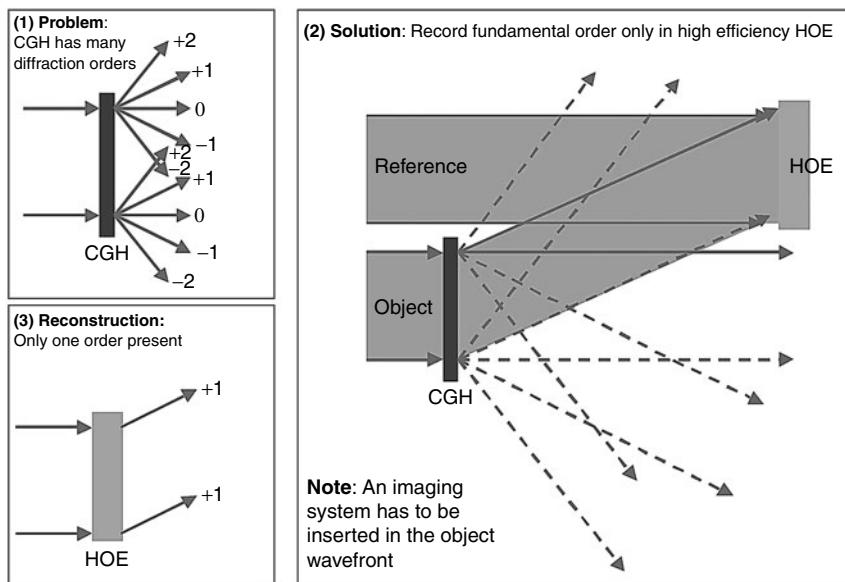


TABLE 13.2 Holographic Materials for the Recording of HOEs

Material	Hologram Type	Application	Advantage	Inconvenient
Photo-resist	Thin (surface relief)	Diffractive optics / lithographic patterning	Simple process, resist pattern can be etched by RIE in substrate	Low efficiency due to resist modulation.
Silver halide	Volume	First holograms	Cheap	Thin, low selectivity
DCG (Dichromated gelatin)	Thick volume hologram	Selectivity applications (optical storage, DWDM demux)	Cheap, relatively thick for strong angular and spectral selectivity	Long term stability
Photopolymer (DuPont type)	Thin hologram	Replication in mass of anticountefering holograms	Replicable in mass	Low resolution unavailable outside DuPont
Photorefractive crystal	Very thick volume hologram	Laboratory experiments	Erasable/rewritable Very high angular and spectral selectivity	Expensive—polarization issues
AO (acousto-optic) modules	Volume Bragg grating	Display	Fast reconfigurable	Only linear gratings, need piezoelectric transducers
PDLC (polymer dispersed liquid crystals)	Volume hologram	Display, telecom, datacom, storage	High index, high Δn , switchable through gray levels, fast	Long-term stability—polarization issues

Diamond Ruling/Turning

Diamond turning or diamond ruling via a computer controlled high-resolution CNC ruler or lathe can be a very effective way to produce high quality-efficient diffractives in a wide range of materials (plastics, glass, metals, ZnSe, Ge, etc.), see Fig. 13.29. However, there are limited to either one-dimensional (linear gratings) or circularly symmetric elements like on-axis symmetric lenses, with smallest fringes several times greater than the diamond tool. In the triangular or sawtooth grating example, the diamond tool can have the exact shape of the grating groove.

Note that more complex 6-axis CNC lathes can generate complex anamorphic fringes and profiles, as it is also done for their refractive counterparts.

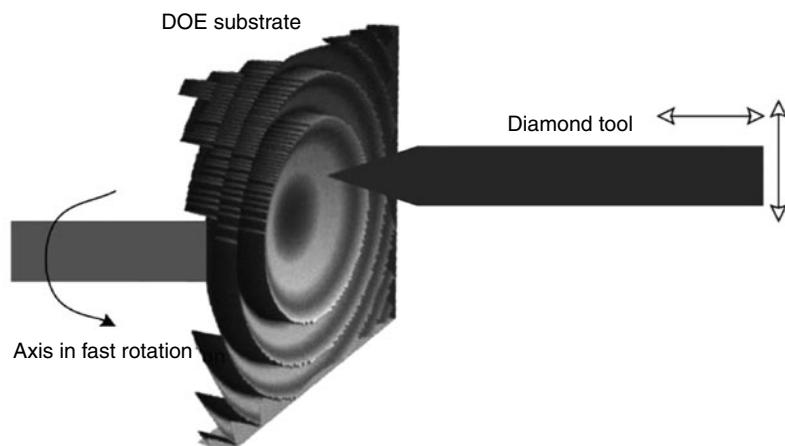
Optical Microlithography

Optical microlithography fabrication technologies and techniques are derived from the standard integrated circuit (IC) industry, and have been applied to the fabrication of diffractives since the mid-1980s.²³

These techniques are used to fabricate diffractive optics when either the following conditions cannot be met by the previously described fabrication technologies (diamond turning and holographic exposure)²⁴:

Figure 13.29

Diamond Turning of a Spherical Blazed Fresnel Lens

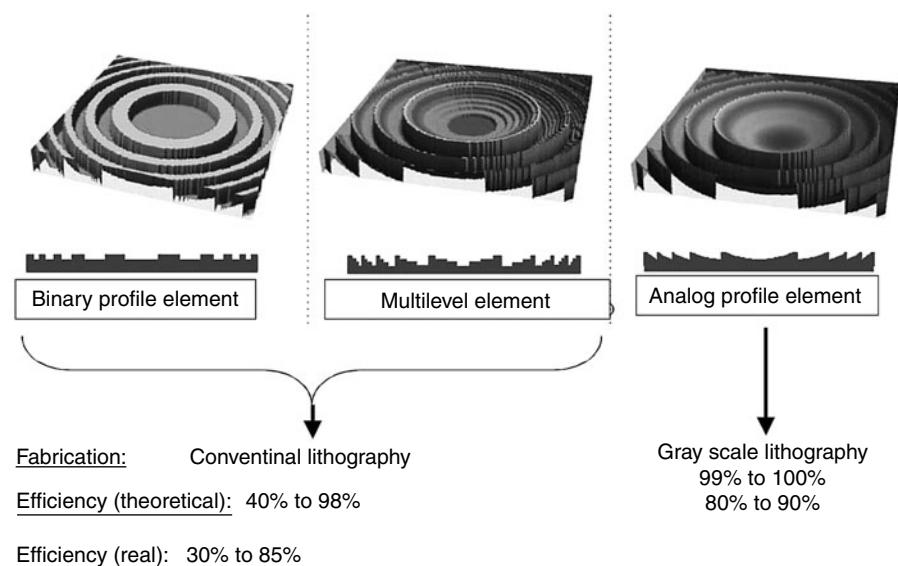


1. Fabrication of nonsymmetric or nonlinear features (CGH-like features)
2. Complex three-dimensional structures which cannot result from holographic interference
3. Precise spatial multiplexing of various diffractives (planar optical bench, arrays)
4. Integration of precise digital micro structures with additional alignment or integration features (fiducials, marks, targets, etc.)
5. Use of durable materials like quartz, glass, silicon, sapphire, ZnSe, Ge, etc.
6. Need for mass replication of the master element in these same materials

Microlithographic fabrication techniques can result in a wide variety of microstructured phase relief elements as depicted in Fig. 13.30 (from binary to multilevel to quasi analog surface relief).

In many cases, the diffraction efficiency is the most important criteria to consider, and therefore a multilevel or quasi-analog surface relief is often desired rather than a simple binary element, but this would also require a high fabrication budget.

Figure 13.30
The Various Microlithographic Techniques used for Diffractives Fabrication



The microlithography fabrication process can be quite cumbersome when the optical designer does not have the adequate CAD tools to generate the required mask layouts.

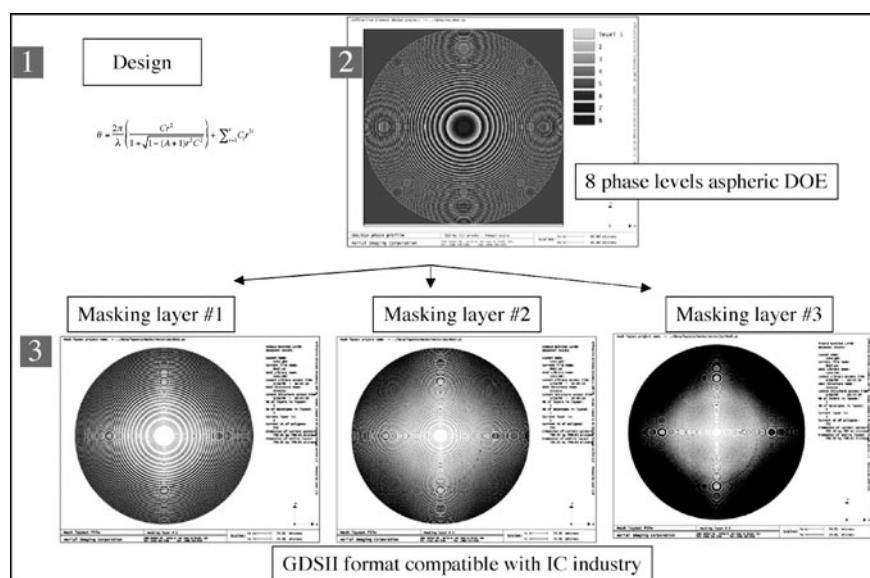
The lithographic fabrication process can be split into five major tasks:

1. Generation of the phase profile (via analytic or numeric techniques)
2. Fracture of the resulting fringes into GDS2 polygons (IC industry standard format for mask layout)
3. Photomask patterning via laser or e-beam patterning system
4. Optical lithography, etching, and dicing
5. Potential replication of master elements by embossing or injection molding

Figure 13.31 shows the first two processes (generation of the layout) for an eight-phase levels diffractive lens for which three successive binary masks are generated and which fringes are fractures into polygons. These three masks are then used in multilevel lithography to generate the eight-phase levels.

Figure 13.31

Generation of the GDS2 Layouts (Successive Masks) from the Initial Lens Description



Multilevel lithographic fabrication is now becoming the most widely used diffractive fabrication technique. Although it is not the ideal one, it is the best understood and a widely available technique.^{22,23} Figure 13.32 shows this process, where 2^N phase relief levels can be fabricated with a set of N binary amplitude photomasks.

There are two different ways to use optical lithography: by contact or projection. Table 13.3 shows the specifications of each lithographic technique, their advantages and limitations. Both are now widely used for diffractive optics fabrication.

EXAMPLE OF MULTILEVEL LITHOGRAPHY FABRICATION

Figure 13.33 show the two successive steps involved in the fabrication of a four-phase levels Fresnel lens (first photo after the first lithography/etching step and second photo after the second lithography/etching step).

Figure 13.32
Conventional Multilevel Lithographic Fabrication Technique

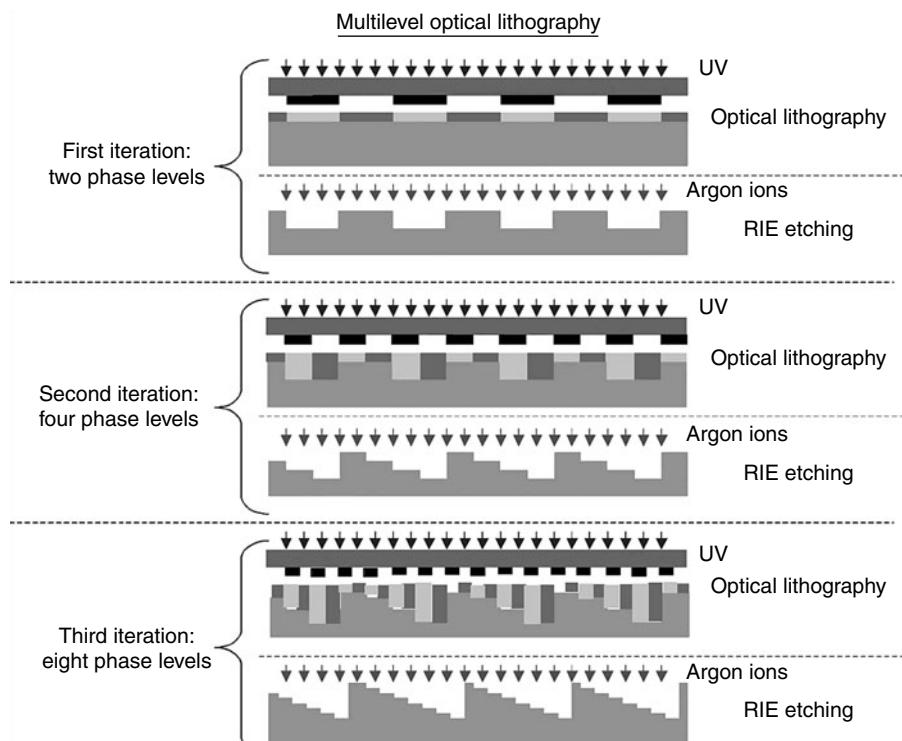
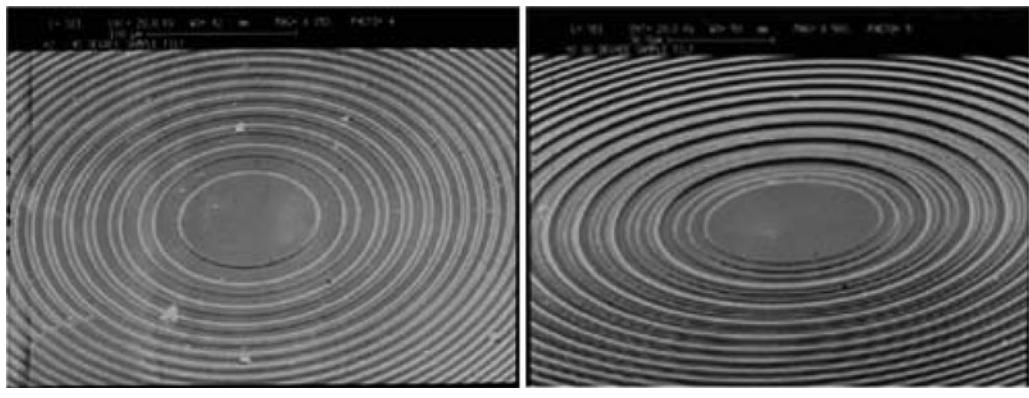


TABLE 13.3

Contact and Projection Lithography for Multilevel Diffractives Fabrication

Parameter	Contact Lithography	Projection Lithography
Lithography tool	Mask aligner	Stepper—Stepper/scanner
Demagnification factor	1×	1× to 10× (mostly 4×)
Minimum feature size printed on wafer	1.0 μm	0.85 μm (h-line) 0.35 μm (i-line) 0.18 μm (deep UV) <0.1 μm (extreme UV or X ray lithography)
Max field size on wafer (for individual element)	Size of the wafer — 0.5"	<30 × 30mm
Minimum lateral field misalignment	Approx 0.5 μm	<0.25 μm
Cost of machine	Medium (250K)	Very high (>5 M)
Costs per batch	Low	High
Minimum batch	Individual wafers	Cassettes batch (25 wafers)
Wafer size versus mask (reticle) size	2"/3" 3"/4" 4"/5" 5"/6"	4"/5" 6"/7" 8"/9" 12"/13"

**Figure 13.33**

Fabrication of a Four-Phase Levels Fresnel Lens with Two Successive Masking Layers

SUCCESSIVE MASK ALIGNMENTS IN MULTILEVEL OPTICAL LITHOGRAPHY FABRICATION In order to align properly the two fields (photomask layers), alignment marks, and alignment fiducials are necessary. The accuracy of the alignment is crucial since any misalignment would decrease the efficiency and would create parasitic diffraction effects²⁵ due to high frequency structures arising from the misregistrations between successive layers (see Fig. 13.34).

Note the spikes generated by the successive field misalignments, which are typically less than 0.5 μm with a contact masking lithography process. The lateral structure sizes are of about 2 μm . The alignment

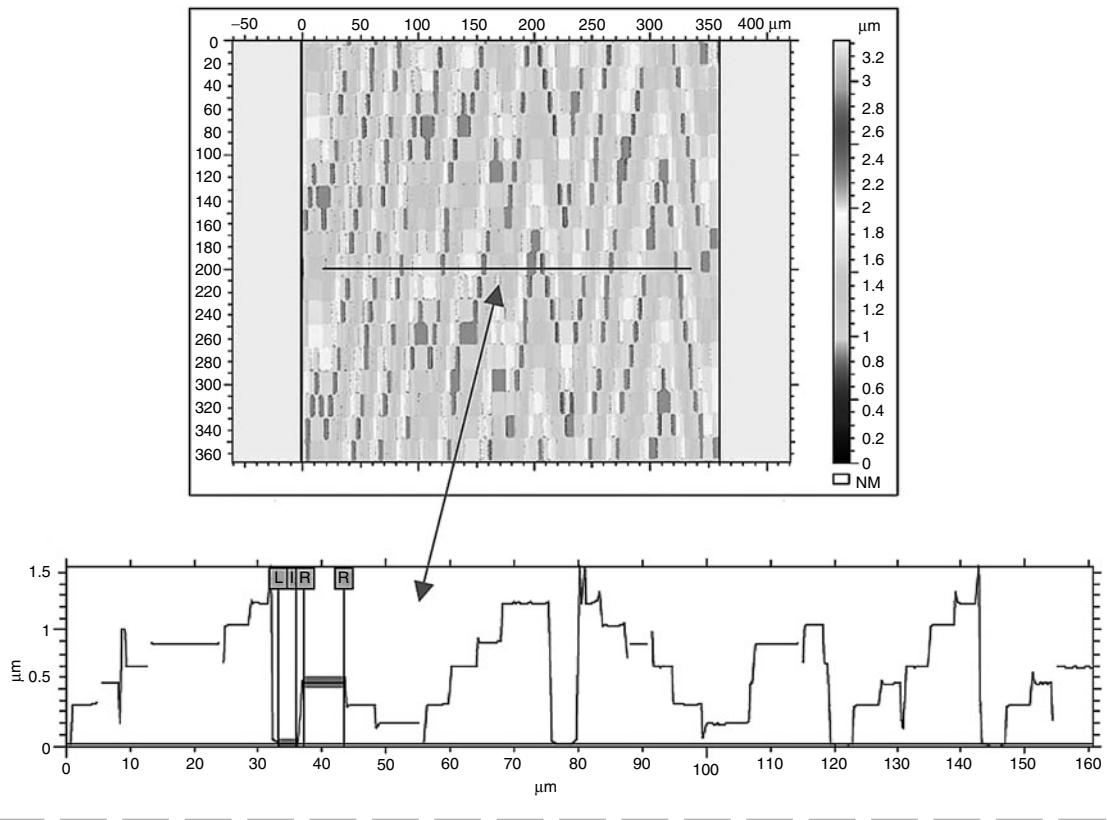


Figure 13.34

Lateral Misregistrations Detected via White Light Interferometry over a Eight-Phase Levels Beam Shaper CGH (Three Successive Masking Layers)

errors here are thus smaller than $0.25 \mu\text{m}$, which nevertheless create high frequency parasitic structures and other artifacts which can decrease the overall performance of the diffractive element.

Figure 13.35 shows a microscope photo of a Fourier type and a Fresnel type CGH fabricated through binary lithography with a single lithography/etching step in quartz.

Note that in the Fourier type CGH no fringe geometry can be seen, as expected and discussed in the previous section. On the other hand, the Fresnel type CGH shows fringe-like features, which can be modeled by the LGA method described in the modeling section.

DIFFRACTION EFFICIENCY CALCULATIONS FOR MULTILEVEL LITHOGRAPHIC FABRICATION

Scalar theory of diffraction can be used effectively to predict the diffraction efficiency of binary or multilevel amplitude or phase elements fabricated by microlithography, in the realm of validity of scalar theory (see also Fig. 13.20).

Amplitude Gratings Efficiency Calculation Based on the Fraunhofer formulation of the diffracted field far away from the diffractive aperture (see modeling section^{3,9}), we will derive here the diffraction efficiency formulation for amplitude gratings (Fig. 13.36).

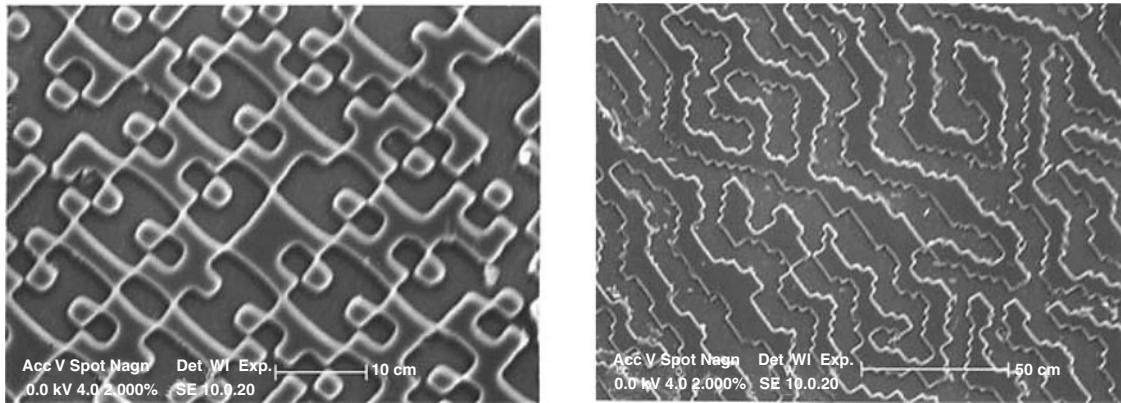
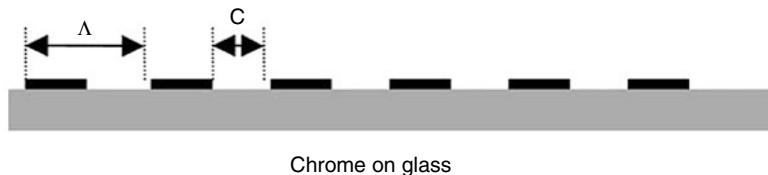


Figure 13.35
Fourier and Fresnel CGH Fabricated in Quartz Substrate as Binary Elements

Figure 13.36
Chrome on Glass
Binary Linear Amplitude Grating



The amplitude of the grating function can be expressed as a Fourier expansion (Eqs. 13.16 and 13.17):

$$a(x,y) = \begin{cases} 1 & \text{if } m\Lambda \leq x \leq c + m\Lambda \\ 0 & \text{elsewhere} \end{cases} \quad (13.16)$$

$$a(x,y) = \frac{c}{\Lambda} \sum_{m=-\infty}^{+\infty} e^{-im(\frac{mc}{\Lambda})} \frac{\sin(m\pi\frac{c}{\Lambda})}{m\pi\frac{c}{\Lambda}} e^{2i\pi\frac{x}{\Lambda}} \quad (13.17)$$

where each exponential term represents a plane wave, and m the index of the different diffraction orders present.

The magnitude of this m th diffraction order can be expressed as:

$$\eta_m = \left| \frac{\sin(m\pi\frac{c}{\Lambda})}{m\pi} \right|^2 \quad (13.18)$$

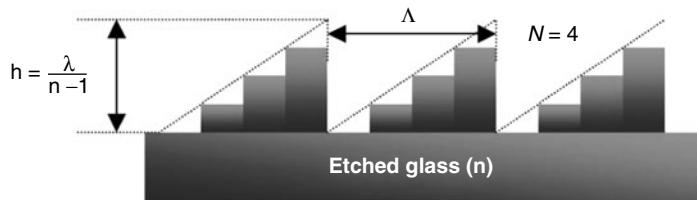
For regularly spaced opaque/transparent lines, (that is, a grating with a duty cycle of 50%), the diffraction efficiency formulation in the m th order becomes (Eq. 13.19):

$$\eta_m = \left| \frac{\sin(\frac{m\pi}{2})}{m\pi} \right|^2 \quad (13.19)$$

Multilevel Phase Gratings Efficiency Calculations Similarly to amplitude gratings, we can derive the diffraction efficiency for multilevel phase relief gratings (see Fig. 13.37).

If the grating is etched so that the maximum optical path difference yields 2π in transmission, which maximizes the efficiency (see introductory

Figure 13.37
Quadratic Phase Surface Relief Grating



section of this chapter), we get the following expression for the diffraction efficiency in the m th order for N phase levels (Eq. 13.20):

$$\eta_m^N = \left| \frac{\sin\left(\frac{m\pi}{N}\right)}{\frac{m\pi}{N}} \right|^2 \quad (13.20)$$

See also Table 13.4 in next section for the predicted values of diffraction efficiencies for multilevel phase relief diffractives.

We can easily extrapolate these results from infinitely long linear gratings to more complex structures limited in space, like Fresnel lenses and general CGHs (which can be approximated locally by such linear gratings).

When the number of phase level grows to 16 or more, the diffractive element can be easily considered as a quasi-analog surface relief element. The diffraction efficiency of an analog surface relief element (for $m = 1$ and $n \rightarrow \infty$) is shown in Fig. 13.38.

In fact, it does not make sense to fabricate a diffractive element with more than 16 levels by conventional multilevel masking techniques, since the successive systematic lateral misalignment errors and cascaded etching depth errors would reduce the diffraction efficiency dramatically. The diffraction efficiency increase when growing from 16 levels to more levels (32, 64, 128, or 256), even for perfect fabrication (which is an impossible task), is infinitesimal (a fraction of a percent), but the fabrication effort is enormous and does not make sense practically.

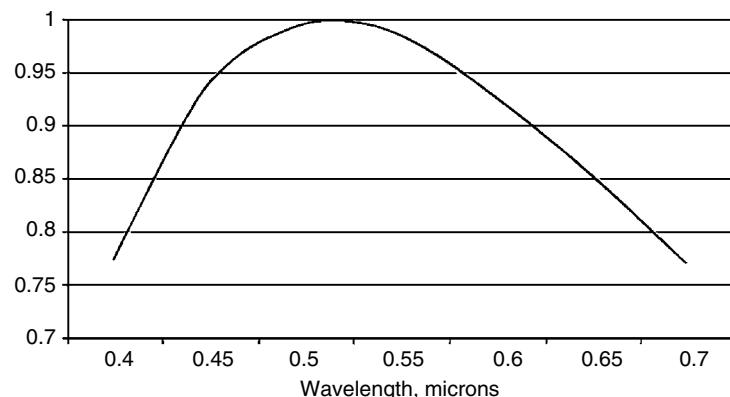
Figure 13.39 shows the diffraction efficiency of a blazed grating (or lens) for the diffraction orders +1 (fundamental), +2 and +3, as a function of the reconstruction wavelength.

The blazed lens in Fig. 13.39 is designed and fabricated to yield maximum efficiency in the fundamental order in the green region of the spectrum (550 nm). As the efficiency in the fundamental drops when going from green to lower wavelengths, the efficiency in the higher orders increases.

TABLE 13.4 The Various Lithographic Fabrication Technologies and Their Characteristics

Microlithographic Technique	Nb of Surface Levels	η (%) Theor.	η (%) Real	Nb of Mask	Litho/Etching Processes	Fabrication Costs	Advantages	Limitations
Binary amplitude	Binary	<8%	<8%	1	0/0	Minimal (only mask)	Cost and resolution	Only one element, very low efficiency
Binary phase in resist	Binary	40.5%	≈30%	1	0/0	Minimal (only mask exposure)	Cost and resolution	Only one element, profile in resist
Direct analog write (laser or e-beam)	Quasi analog	>80%	≈80%	1	0/0	Medium, requires dosage modulation	Direct write with laser or e-beam, no lateral alignment issues	Only one element, profile in resist, difficult to etch
Binary phase in mask substrate	Binary	40.5%	≈35%	1	0/1	Small (only mask and RIE)	Cost and resolution, stable material	Only one element
Multilevel (N masks for 2^N levels)	4	81%	≈75%	2	2/2	Medium	Efficiency Many elements	Mask mis-alignments and etch depths
	8	95%	≈85%	3	3/3	High	High efficiency Many elements	Costs and cascaded misalignments
	16	99%	≈90%	4	4/4	Very high	High efficiency Many elements	Costs and cascaded misalignments
	>16	>99%	≈90%	>4	>4/4	Very high	High efficiency Many elements	Very high costs
Grey scale lithography	Analog	>90%	≈90%	1	1/1	Very high	High efficiency Many elements No misalignment	Costs

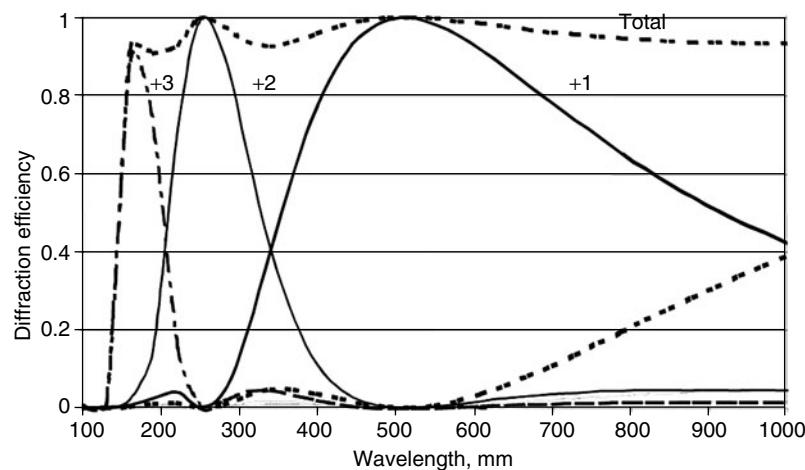
Figure 13.38
Diffraction Efficiency in the Fundamental Order for a Quasi Analog Surface Relief Profile (Lens or Grating). The Optimum Wavelength (Design Wavelength) Is 510 nm

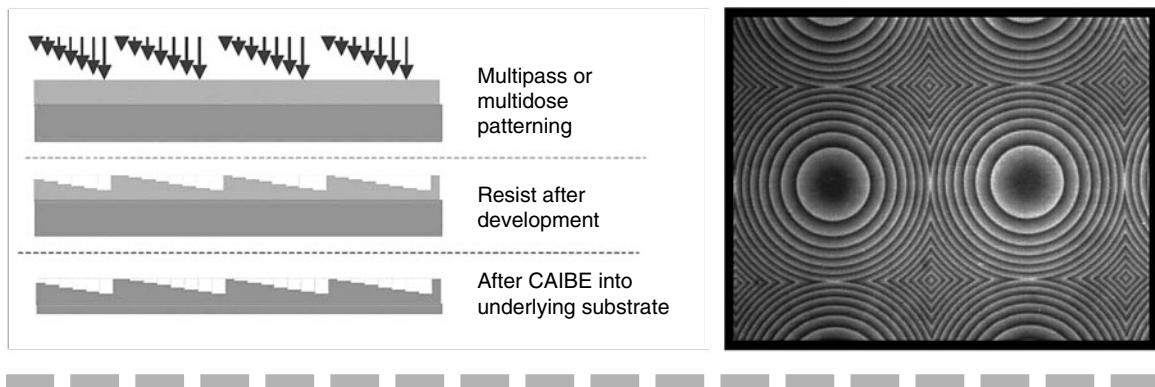


Direct Write Techniques In order to avoid the systematic lateral field misregistration errors occurring when using conventional multilevel lithography, direct write methods have been reported.²⁶ These methods write directly into a resist layer a quasi-continuous phase profile, by either with e-beam or laser beam dosage modulation or multipass patterning (see Fig. 13.40).

GRAY SCALE OPTICAL LITHOGRAPHY Several gray scale lithographic fabrication techniques have been reported since the mid 1990s (HEBS glass—high energy beam sensitive glass or inorganic resist with dry development). They have in common the fabrication of a gray scale mask or reticule used in projection lithography in order to expose

Figure 13.39
Diffraction Efficiency versus Wavelength for the Three Diffraction Orders +1, +2, and +3 for a Blazed Grating



**Figure 13.40**

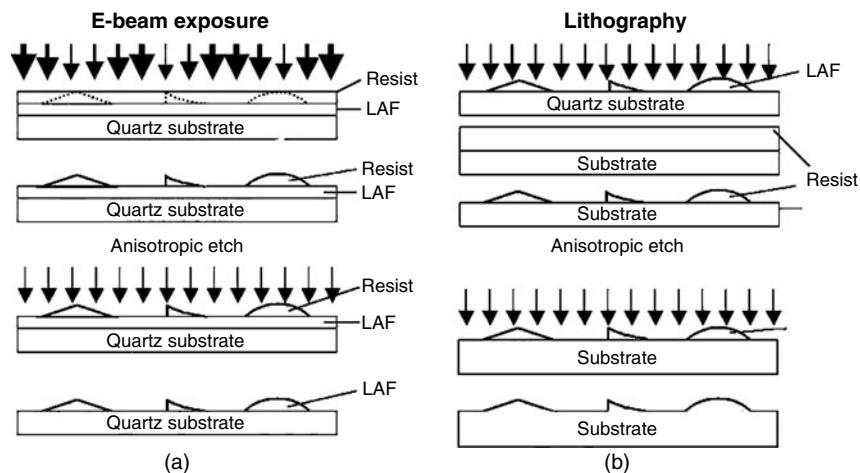
Direct Write Technique and Resulting 256 Levels Fresnel Lens Array Etched into Quartz

directly a photo resist in an analog way, and thus yield an analog surface relief modulation.

Figure 13.41 illustrates the schematic fabrication process of gray-scale technology utilizing a new carbon-based mask material and improved lithographic/etching processes.²⁷

The transmittance of the mask is related to the thickness d of their light attenuating film (LAF) according to $\exp[-\kappa d]$, where κ is the absorption coefficient of the LAF. When the gray scale mask is used in lithography and dry etching processes with special attentions paid to preserving the analog features in the mask, many gray scale device

Figure 13.41
 (a) Fabrication Process of a New Carbon-Based Gray Scale Mask with Conventional E-Beam Resist
 (b) Fabrication of Gray Scale Device Structures Using the Gray Scale Mask, an Analog Resists and an ICP-RIE System (Courtesy of Dr Sing Lee and Dr Jo Zhou, UCSD)



structures can be fabricated without going through the cumbersome lithography and etching processes involving multiple masks. Figure 13.38 shows an AFM picture of an off-axis diffractive lens with a height of about 1 μm and periodicities ranging from 10 to 12 μm , fabricated in quartz by employing the new gray scale mask and 248-nm-stepper lithography.²⁸

Figure 13.42 shows a microscopic picture of a corner cube (with a depth of about 10 μm at its apex) array, fabricated in silicon by the use of the gray scale mask and an i-line aligner.

The gray scale structures have surface smoothness corresponding to 64 thickness or phase levels in Fig. 13.42 (left) and 256 levels in Fig. 13.42 (right).

There have also been pseudo gray scale lithography techniques reported in literature, which use simple binary masks to encode gray levels through either pulse width or pulse density modulation of opening in a chrome layer (see Fig. 13.43).

The gray scale intensity mapping is created when light is diffracted through these small openings on the photomask and then reimaged onto the wafer in the resist layer.

SUMMARY OF FABRICATION TECHNIQUES Table 13.4 shows the various lithographic fabrication technologies used in industry, the resulting typical diffraction efficiencies, as well as their respective advantages and limitations.

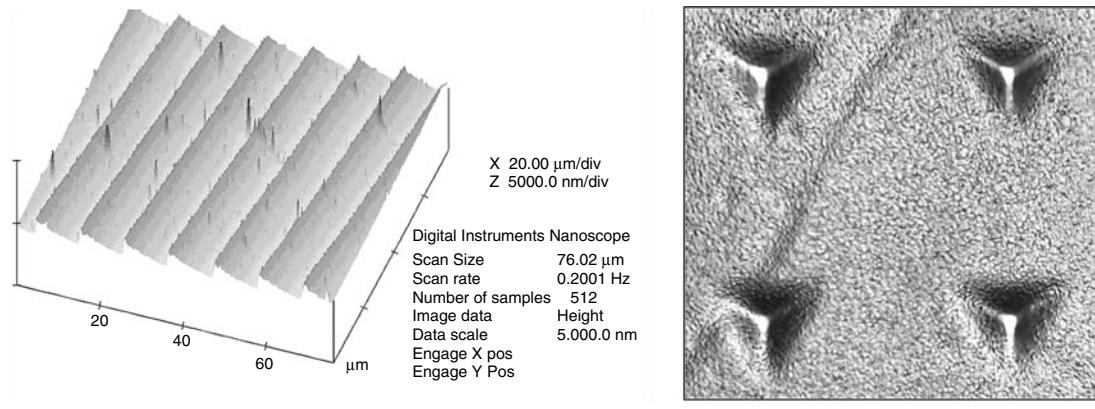
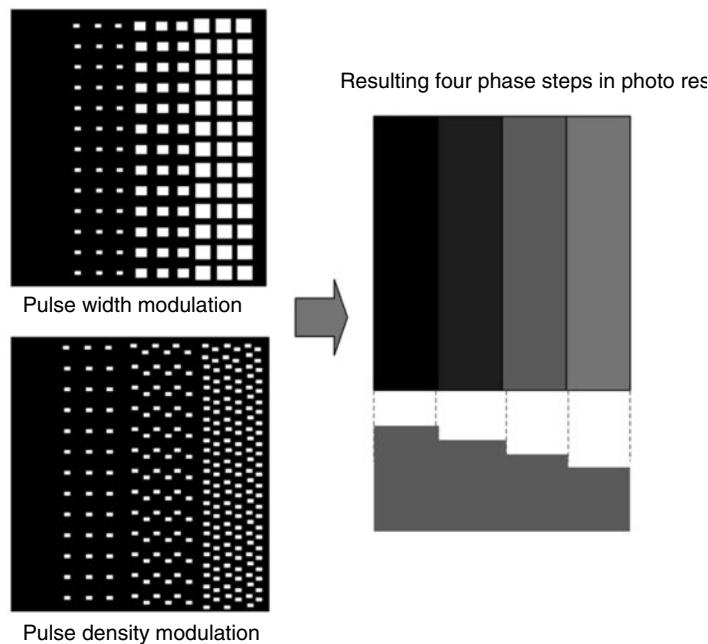


Figure 13.42

Gray Scale Lithography Process—Off-Axis Diffractive Lens on Quartz (Left) and Array of Microcorner Cubes (Right)

Figure 13-43
Pseudo Gray Scale
Binary Amplitude
Masks for the Gen-
eration of Multilevel
Surface Relief by
Pulse Density or Pulse
Width Modulation

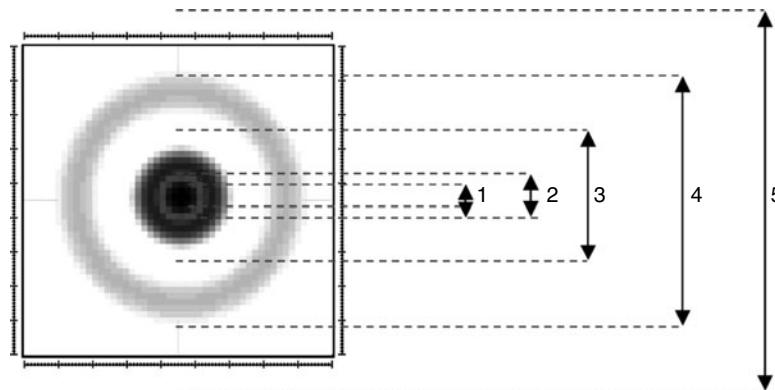


IMPORTANT REMARKS: NOTION OF DIFFRACTION EFFICIENCY FOR INDUSTRIAL APPLICATIONS It is important to remember that diffraction efficiency values are theoretical values for the efficiency in the m th diffraction order (for example, the fundamental positive order where $m = 1$). However, for a given application, the diffraction efficiency is actually very often defined as a fraction of this value or even as a combination of several diffraction orders. In order to illustrate this concept, we show below two examples where the efficiency linked to an application is respectively lower and higher than the theoretical efficiencies as calculated here before.

a. Analog surface relief Fresnel lens for data storage.

In this case, we are not interested in any imaging task from the lens, but rather in the amount of energy within the waist of the focal spot (see Fig. 13.44). The theoretical efficiency relates to the amount of energy in this spot (within the Airy disk), but also the energy within the successive side lobes. Therefore, the efficiency linked to this application is lower than the predicted efficiency of

Figure 13.44
Definition of
Application-Oriented
Diffraction Efficiency



the element. The figure shows the focal spot and five different definitions of efficiency in this case (1: FWHM, 2: waist, 3: Airy disk, 4: First side lobe, 5: etc.).

b. Binary Fourier CGH with on-axis symmetric reconstruction.

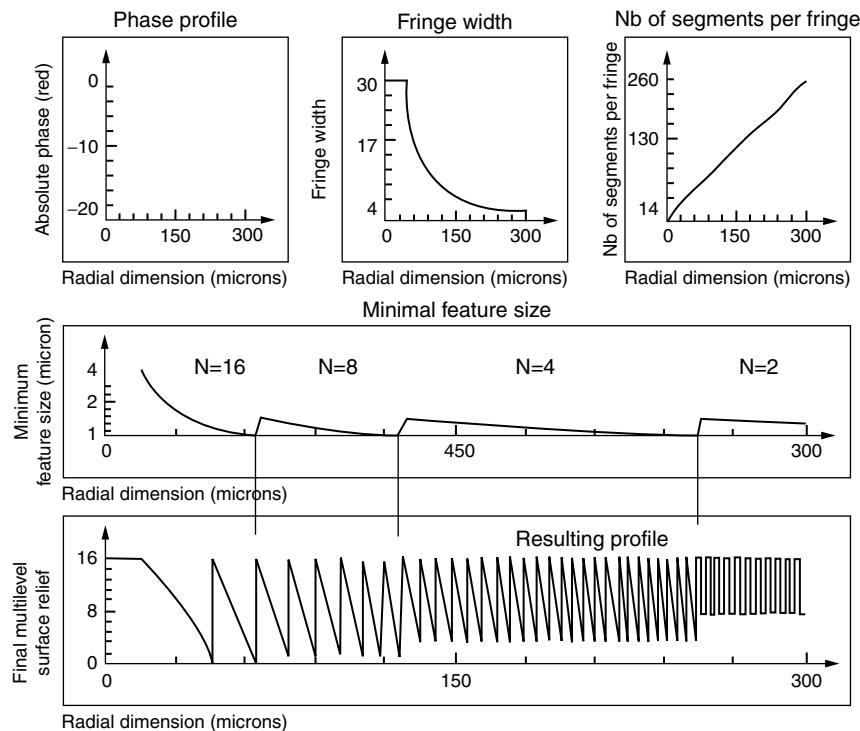
In this case, we are taking advantage of the fact that in the Fourier plane of such an element, two main reconstructions appear on-axis with the same efficiency (about 40%). These are the two conjugate fundamental orders (orders +1 and -1). If the reconstruction is symmetric in the Fourier plane, these two reconstructions can overlap exactly and double the theoretical efficiency from 40.5 to 81%.

This technique is often used for the binary (that is, relatively cheap) fabrication of high-efficiency on-axis spot array generators, logo generators, beam shapers, etc.

EFFICIENCY CALCULATIONS OF MULTILEVEL FRESNEL LENSES IN REAL WORLD In many cases, the microlithographic fabrication limitation is the lateral size of the smallest structure (note that a similar limitation also applies to the validity of the scalar theory of diffraction). As the fringe width (local period) of a Fresnel lens decreases radially, the maximum number of phase levels approximating the analog fringe relief surface also decrease. Therefore, a smart way to fabricate an efficient lens considering the limitation of the fabrication tool is to optimize the number of phase levels over each fringe. In such a lens, the efficiency is maximal in the center, and decreases radially according to the number of phase levels present (see Fig. 13.45).

Figure 13.45

Optimum Fabrication
of a Multilevel Fresnel
Lens for a Specific
Fabrication
Resolution



FINAL WAFER DICING Finally, AR coating and wafer dicing concludes usually the fabrication of diffractives through microlithography. Figure 13.46 shows a 4-in. quartz wafer AR coated and diced into individual rectangular dies (still held on sticky blue tape) and ready to be inserted in their respective products.

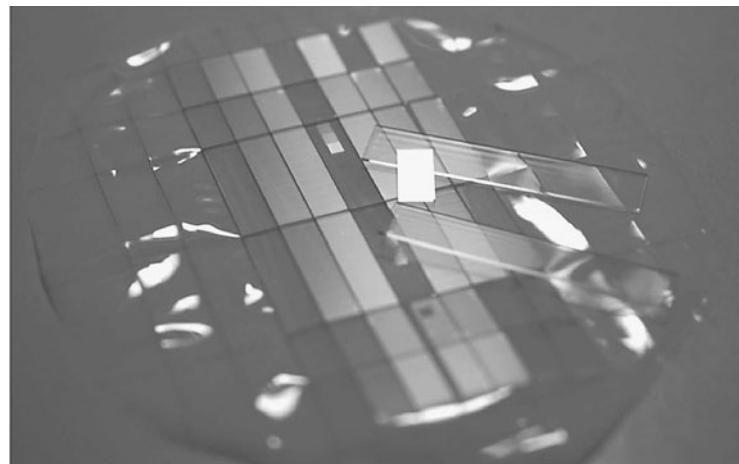
Mass Replication of Diffractives

The fabrication of diffractive optics makes sense in an industrial realm only if these elements can be cheaply replicated in mass. This is the *De Facto* condition for diffractive optics to emerge from the lab and address industrial needs in consumer electronics products and other application fields like automotive, factory automation, biomedical, and telecom, etc.

In the previous sections, we have emphasized the generation of master elements, either by holographic recording, diamond machining, or by microlithography.

Figure 13.46

Diced and AR Coated Quartz Wafer



There are several methods available to the optical engineer to replicate these master elements in volume. These include roll embossing, hot embossing, UV casting, injection molding, and Sol-Gel process (see Fig. 13.47).

Mass replication of a positive master element includes three successive processes:

- Generation of positive master through microlithography, diamond turning, etc.
- Nickel electroplating of negative master (negative structures on Nickel film)
- Generation of a final mold and replication (embossing or injection molding)

Figure 13.47

Mass Replication Process by Embossing or Injection Molding

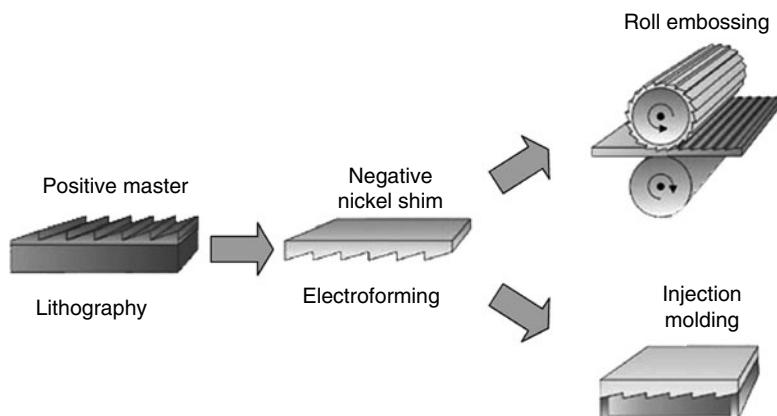


Figure 13.48 shows photos of diffractive optics replicated by CD-ROM injection molding, which is a very effective way to produce in mass diffractives on a polycarbonate substrate, the same thickness and size as a standard CD-ROM.

Table 13.5 summarizes the main diffractive optics replication technologies used in industry today.

Such replication technologies, which for most of them have been developed and optimized for other technologies like CD and DVD disk replication, provide an invaluable asset to diffractive optics to finally come out of the academic/research arena, and propose concrete industrial solutions to real market needs in many consumer electronics, factory automation and telecom applications, as we will see in the next section.

Where Are Diffractives Used?

Diffractive optics is used in numerous applications today, ranging from consumer electronics, sensors, factory automation, to telecom and biomedical applications.

Spectroscopic Applications

Historically, spectroscopy has been the first great application realm for diffractives, and more precisely for linear ruled gratings (see also Figure 13.7 in first section of this chapter). For many people, including many optical

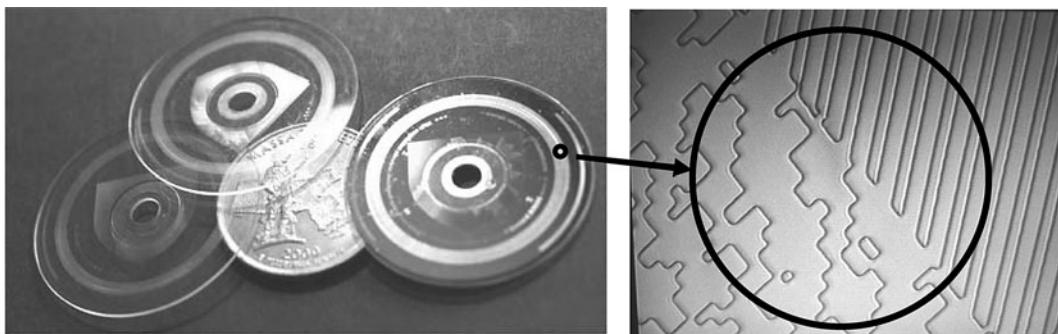


Figure 13.48

Replication of Diffractive Encoder Structures by CD Injection Molding

TABLE 13.5 Summary of the Major Diffractive Optics Replication Technologies Used in Industry Today

Replication Technology	Element Type	Typical lot Size	Price per Replica	Resulting Substrate	Advantages	Limitations
Conventional optical lithography	Binary/multilevel surface relief	25 wafer cassette lot	High (\$0.1K–\$1K)	Any wafer (quartz, fused silica, glass, silicon, etc.)	Very high precision via optical lithography and dry etching	Pricing, lot size, dicing
Grey scale lithography	Analog surface relief	Individual wafers	Very costly (>\$1K)	Selected wafers	Very high efficiency	Costs and availability
CD/DVD injection molding	Any surface relief	>1000s to Millions	Cheap <\$1	Polycarbonate	Rigid substrate, proven technology	Need for specialized equipment and Nickel shim generation
Planar hot embossing	Any surface relief	1000s	Cheap \$1–\$10	Any plastic	Rigid substrate, no need for specialized equipment	Nickel shim generation needed
Roll embossing	Any surface relief	Millions	Very cheap <\$0.1	Mylar, thin plastics,...	Cheap proven technology	Substrate not rigid, Ni shim generation
UV casting	Any surface relief	1000s	Cheap \$0.1 to \$10	Polymer	No need for special equipment	No rigid substrate, shrinkage, tear, wear
Sol gel	Any surface relief	100s	Average \$10–\$100	Glass	Resulting element is silica based (glass)	Shrinkage and geometrical control

engineers, spectroscopic gratings remain the one and only application of diffractive optics today, a notion that has to be revised as we will see in the other sections.

Imaging Applications

Incorporating a diffractive lens in an imaging system is usually considered difficult when dealing with broadband illumination. However, as seen in Fig. 13.11 in this chapter, diffractives can help reduce chromatic and thermal aberrations and reduce the overall number of lenses needed. Canon Inc. of Japan (Fig. 13.49) has introduced this concept in several telefocus objective lenses for their line of digital single-lens reflex (SLR) cameras. Canon uses a set of sandwich Fresnel lenses with different refractive indexes in order to reduce the overall size and weight of the objective. Not only is the resulting objective shorter, it also incorporates lesser lenses.

In a general way, it is much simpler to design a hybrid achromat for single wavelength, for example, infrared cameras for automotive three-dimensional sensing in bumpers, camera lenses for laser-based factory automation sensors, etc.

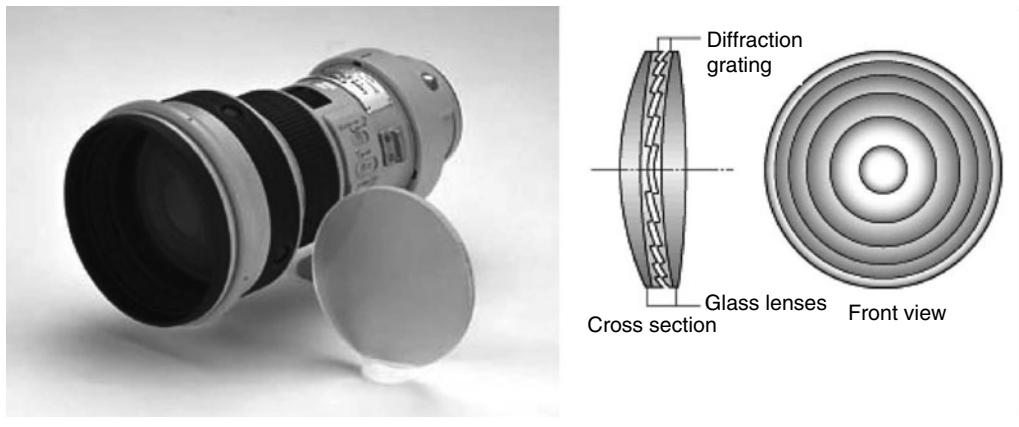


Figure 13.49

Canon Ltd. Diffractive Tele-Objective Lenses

Optical Computing and Optical Interconnections

Optical computing has become, through the 1990s a very hot topic of research and development, and has extensively used diffractive optics for both optical clock distribution²⁹ in multichip modules (MCM) and optical interconnections in massively parallel computing architectures.³⁰ However, this scheme has never made it to a consumer electronic reality, and remains a development area mainly for high-demanding military computing applications.

Optical Data Storage

Optical data storage has become the first consumer electronics application realm for diffractive optics. Actually, many optical functionalities in a standard OPU unit for CD/DVD drives incorporate diffractives today [beam splitting for tracking purposes and hybrid refractive/diffractive lenses for dual spot lens for CD/DVD read-out (see section "How Are Diffractives Fabricated?" for more insight into this hybrid lens architecture and for more recent all diffractive data storage applications see www.inphase-technologies.com]. Figure 13.50 shows fast diffractive lenses used in Winchester heads for magnetic optical disk read-out (left), and dual focus lens for CD/DVD read-out.

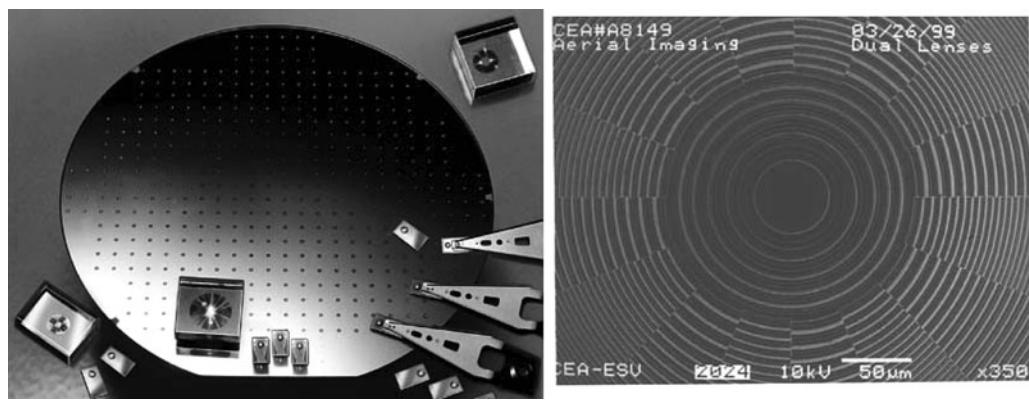


Figure 13.50

Fast Lenses for Magneto Optical Winchester Flying Heads and Dual Focus Lenses for CD/DVD Read-Out

Optical Telecom

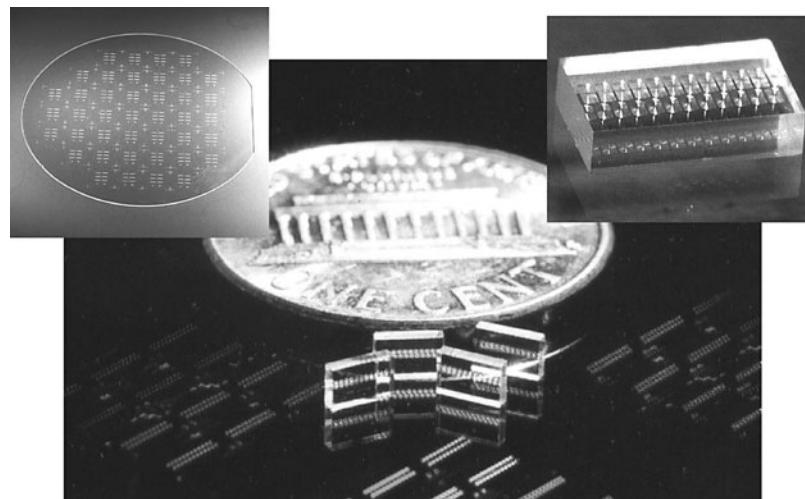
Optical telecom has seen massive surge of interest after the DWDM revolution at the end of the last decade, and more recently with the steady growth of 10-Gbs optical Ethernet lines. Diffractives have mainly been applied (as for spectroscopic gratings) to spectral Demux and Mux applications, as linear ruled gratings (www.jdsu.com) or phase gratings (www.bayspec.com). These Mux assemblies can then be applied to more complex functionalities like optical add-drop modules.

Another realm of application of diffractive in telecom is the introduction of Bragg gratings for very selective spectral filter implementation, both in Bragg reflection and Bragg coupling regimes.

More recently, diffractives have been applied to 10-Gbs optical Ethernet lines for fiber coupling and detector focusing, signal and other monitoring functionalities (www.doc.com). Figure 13.51 shows a 12 lines 10-Gbs optical Ethernet optical assembly block for 850-nm VCSEL laser to fiber coupling and fiber to detector using dual side substrate patterning.

In Fig. 13.51, the left part shows a 6-in. quartz wafer with many individual optical block assemblies, and the right part (as well as the centre parts) show a single diced out assembly ready to be integrated in a 12 array 10-Gbs Ethernet fibre bundle.

Figure 13.51
Multichannel 10Gbs
Ethernet Optical
Block Assembly for
850-nm VCSEL
Lasers/Detectors



Optical Anticounterfeiting

Holograms have been used since the 1970s to implement anticounterfeiting devices in products like credit cards, bank notes, passports, and other critical documents for authentication purposes. Recently, optical variable devices (OVDs), which are mainly spatially multiplexed gratings and lenses, provide additional security to such devices (www.ovd-kinegram.com). These OVDs and other synthetic holograms are for most of them not holographically generated, but rather generated by computer as series of DOEs or CGHs, mastered by conventional optical lithography and replicated by roll embossing. The current trend is to include in such visual OVDs additional machine readable (laser scanner readable) features which can include large amounts of digital data, much larger than the data that can be integrated in radiofrequency identification (RFID) devices. Of course, these new hybrid visual/machine readable diffractive security tags remain read-only, but are very cheap to replicate in mass by roll embossing.

Laser Material Processing (Cutting, Welding, Engraving)

Laser cutting, laser welding, and laser engraving are very desirable functionalities for the laser material processing industry, a market that is growing at a steady rate, since more than two decades. Laser beam shaping for high power CO₂ and YAG (yttrium aluminum garnet) lasers are now common elements, usually fabricated by diamond turning in ZnSe, ZnS, or Ge substrates.

More complex CGHs, usually fabricated as reflective elements etched in quartz with gold coatings, incorporating specific beam engineering functions (like complex logo engraving or accurate intensity redistribution for welding applications) show to be promising solutions for fast but accurate laser material processing tasks, without having any moving part in the optical path (no moving mirrors, shutters, etc.).

Industrial Optical Sensors Applications

Industrial optical sensors have been a large pool of application for diffractive optics. Gas sensors, position, displacement and motion sensors, strain, torque and force sensors, spectral sensors and index of refraction

sensors, Doppler sensors as well as diffractive microsensors in micro-electromechanical systems (MEMS) configuration have been reported in literature and industry (see for example www.microesys.com and www.appliedopto.com). Industrial optical sensors (especially motion encoders) are based on a large and steady market that is growing slowly and has not been affected by technological bubbles as the DWDM optical telecom, optical data storage, or biomedical markets.

Diffractive optics is being used increasingly in infrared image sensors for remote three-dimensional sensing in automotive and general factory automation. A CGH (type 3 element) is designed to project in the far field a set of grids through an IR laser diode. This scene is then acquired by a digital camera with an IR filter which analyzes the deformation of the IR lines and computes back the three-dimensional topology of the scene.

Biomedical Applications

More recently, biomedical applications have been showing great interest in diffractive optics, for specific applications in optical sensors, genomics, proteomics as well as individual cell processing.

Diffractive optics has been used to implement surface plasmon sensors and index sensors by using the Bragg coupling effect in integrated waveguides. Assay analysis in genomics is using CGHs as laser beam splitters in order to illuminate large numbers of individual arrays of samples and read out fluorescence measurements. Laser tweezers have use the specific advantages of dynamic diffractives in a microscope column in order to steer in real time individual cells in a two-dimensional pattern.

More generally, diffractives can be used in any biomedical apparatus that requires homogenization and/or shaping of the Gaussian laser beam into a specific intensity mapping (for example, fluorescence measurements in hematology by uniform laser illumination on blood sample).

Consumer Electronics Applications

When a technology makes it to standard consumer electronics devices, this technology enters the realm of mainstream technology. Although there has been several introductions of diffractive to consumer electronics as described in the previous optical data storage section, no other application has shown to the public the direct and real potential

of diffractives (and especially CGHs) more than the virtual keyboard application (www.canesta.com), the auto-focus pattern generator application (www.sony.co.jp), and of course the laser pointer pattern generator which has been sold in millions since its debuts in 1998 (see Fig. 13.52).

Projection Display Applications

The market expectations for LED and laser based projection displays, both in front (conventional projectors and pocket projectors) and in rear projection (RPTV) architectures, are very promising since current technologies like plasma screens, LCD and regular high pressure arc lamp projection engines are becoming obsolete, as both display screens get larger, projection engines get smaller and prices are falling.

Diffractive optics have many desirable facets which bring elegant solutions to many of the optical tasks involved in future projection display especially as they will be laser based (vertical cavity surface emitting laser [VCSEL] et diodes). For example, CGHs can implement beam shaping, diffusing, homogenizing, steering and despeckeling, as well as direct far field image pattern generation functionalities. One of the first diffractive image pattern generators is presented in Fig. 13.53 (www.light-blueoptics.com, see also www.holoeye.com). Type 5 dynamic diffractive elements are used to implement such functionality (see also Fig. 13.6).

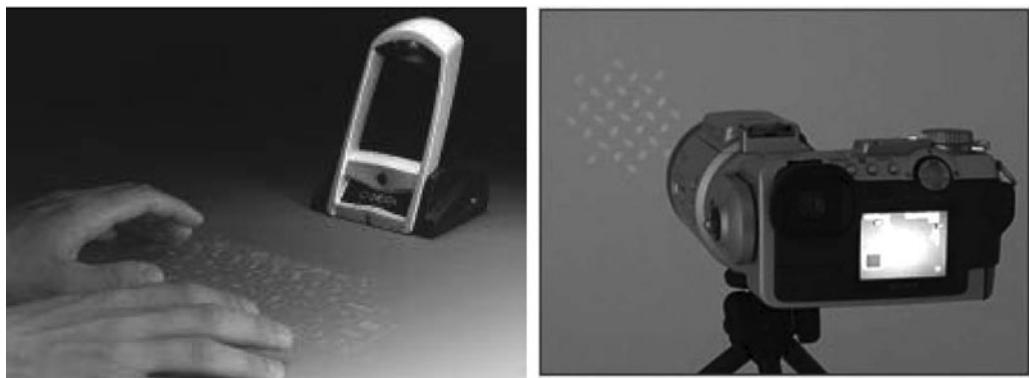
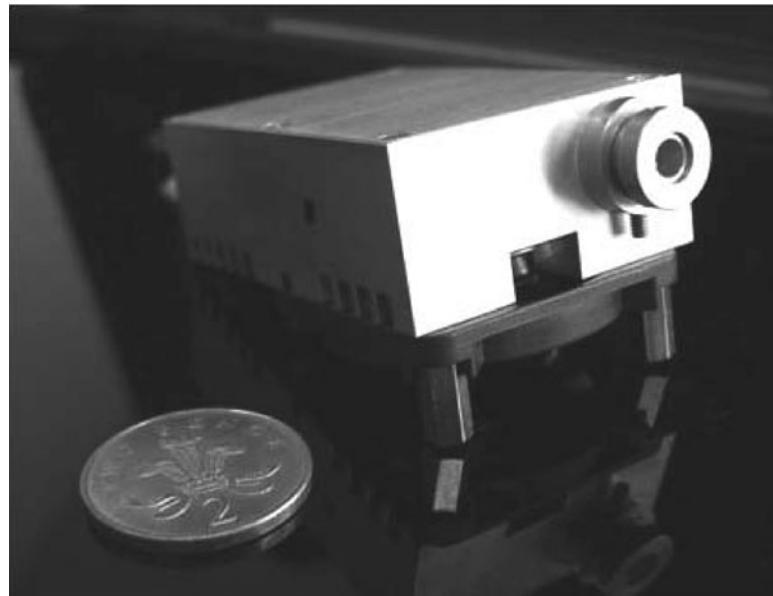


Figure 13.52

Diffractive Virtual Keyboard (Left) and Diffractive Focus Measurement Elements (Fourier Pattern Generators) for Digital Auto-Focus Camera (Right)

Figure 13.53
Diffractive Miniature Projector from Light Blue Optics



Since such a projector generates a diffracted image in the far field, there is no need for an objective lens and/or a zoom lens, and thus the size of the projector can remain very small. However, due to small diffraction angles generated by dynamical diffractives, an additional optical setup is usually necessary to increase these angles.

Niche Markets

Interests for diffractive optics for small niche markets have been demonstrated in the past years, and include complex illumination generation for deep UV stepper lithography for RET applications (reticule enhancement technology)—see www.doc.com.

Market Analysis and Future Applications

Technological waves and investment hypes have historically generated regained interests in diffractive optics technology during the last three decades. These successive technological waves can be described along seven main lines (see Fig. 13.54).

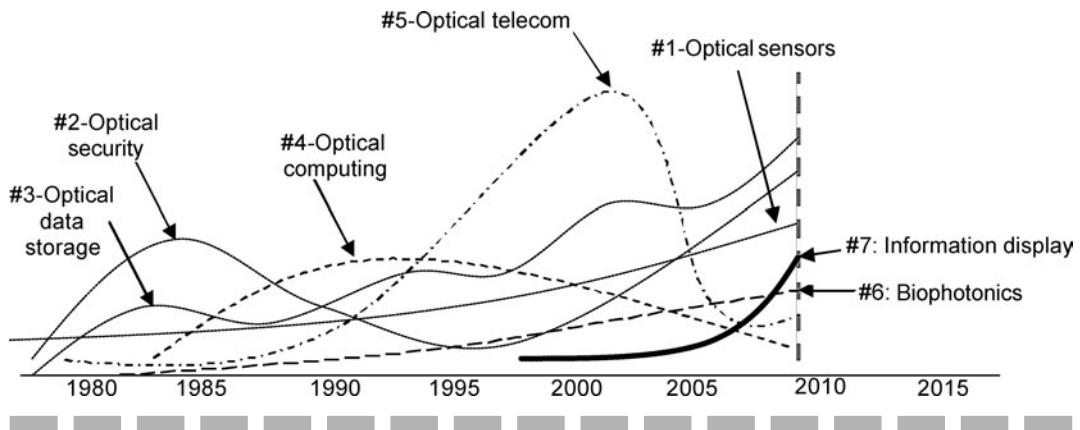


Figure 13.54

Optical Technological Market Waves that Have Fueled Interest Diffractive Optics Technology over the Past 30 Years

- Optical security devices and optical data storage have been the first ones to generate interest in diffractives (other than spectroscopy applications). Optical data storage is actually generating successive interests in diffractives, from the 1980s for CD signal tracking, to the 1990s for DVD read-out heads, and to today's Blu Ray technology and holographic storage.
- Optical computing has generated a lot of interest in diffractives in the early 1990s, but has never been able to bring diffractives to mainstream market (see previous section on optical computing).
- The investment hype of the late 1990s in optical telecom and especially in DWDM Mux Demux devices has ill-fueled interest in diffractives, with the consequence of a disastrous bubble burst in 2003. Recent regain interest in 10Gbs optical Ethernet is correcting this market trajectory (see previous optical telecom section).
- Optical industrial sensors have shown the most steady growth during the past decades, and are still one of the most desirable and less risky markets for the implementation of diffractive optical technology.

- Biophotonics has shown much interest in diffractive very recently, and this interest is growing fast (see previous section on biomedical applications).
- The current interest in information display, and especially in laser-based projection displays (rear and front projection engines), have the potential to build a large market and stable application pool for the use diffractive optics for the next decade.

References

1. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, Oxford, 1993.
2. A. W. Lohmann and D. Paris, Binary Fraunhofer Holograms generated by computer, *Appl. Opt.*, 6 (10), 1739–1748, 1967.
3. B. Kress and P. Meyrueis, *Digital Diffractive Optics*, 1st ed, John Wiley and Sons, Chichester, 1999.
4. G. J. Swanson, *Binary Optics Technology: Theoretical limits on the Diffraction Efficiency of Multilevel Diffractive Optical elements*, MIT Laboratory Report 854, MIT, Cambridge, MA, 1991.
5. D. Leseberg et al., Computer generated holograms: Cylindrical, conical and helical waves, *Appl. Opt.*, 26 (20), 4385–4390, 1987.
6. T. Stone et al., Hybrid diffractive/refractive lenses and achromats, *Appl. Opt.*, 27 (14), 2960–2971, 1988.
7. C. W. Londono et al., Hybrid diffractive/refractive lenses and achromats, *Appl. Opt.*, 27 (14), 2960–2971, 1988.
8. G. P. Berhmann et al., Color correction in athermalized hybrid lenses, *OSA Tech. Dig.*, 9, 67–70, 1993.
9. J. W. Goodman, *Introduction to Fourier Optics*, 2d ed. McGraw-Hill, San Francisco, 1996.
10. J. R. Fienup, Iterative Method applied to image reconstruction and to computer generated holograms, *Opt. Eng.*, 19, 297–305, 1980.
11. K. H. Hoffmann and P. Salomon, The optimal simulated annealing schedule for a simple model, *J. Phys. A*, 23, 3511–3523, 1990.
12. U Malhab et al., Iterative Optimization Algorithms for filter generation in optical correlators: a comparison, *Appl. Opt.*, 31 (8), 117–1125, 1992.

13. R. W. Gerchberg and W. O. Saxton, A practical algorithm for the determination of phase from image and diffraction plane pictures, *Optik*, 35, 237–246, 1972.
14. H. Farhosh et al., Comparison of binary encoding schemes for electron-beam fabrication of computer generated holograms, *Appl. Opt.*, 26, 4361–4372, 1987.
15. B. K. Jennison et al., Iterative approaches to computer generated holography, *Opt. Eng.*, 28 (6), 629–637, 1989.
16. R. Petit, Ed., Electromagnetic theory of gratings, in: *Optics in Current Physics*, Vol. 22, Springer Verlag, Heidelberg, 1980.
17. J. N.Mait and D. W. Prather, Eds., *Selected Papers on Subwavelength Diffractive Optics*, SPIE Milestone Series, Vol. M2 166. SPIE Press, Bellingham, WA, 2001.
18. W. C. Sweatt, Mathematical equivalence between a holographic optical element and an ultra high index lens, *J.Opt. Soc. Am. A*, 69, 486–487, 1979.
19. P. P. Clark et al., Ray tracing models for diffractive elements, *Opt. Soc. Am. Tech. Dig.*, 8, 2–3, 1993.
20. T. K. Gaylor and M. G. Moharam, Analysis and applications of optical diffraction by gratings, *Proc. IEEE*, 73, 894–937, 1985.
21. T. K. Gaylor and M. G. Moharam, Planar Dielectric grating diffraction theories, *Appl. Phys. B*, 28, 1–14, 1982.
22. W. B. Veldkamp, Binary Optics: The optics technology of the decade, in: *Proceedings of 37th International Symposium on Electron, Ion and Photon Beams*, San Diego, CA, 1993.
23. M. W. Farn, Binary Gratings with increase efficiency, *Appl. Opt.*, 31, 4453–4458, 1992.
24. Vasara et al., Binary surface relief gratings for array illumination in digital optics, *Appl. Opt.*, 31, 3320–3336, 1992.
25. M. W. Farn et al., Process error limitations on binary optics performance, in computer and optically generated holographic optics, *SPIE Proc.*, 1555, 80–88, 1991.
26. K. S. Urquhart and S. H. Lee, Computer Generated Holograms fabricated by direct write of positive electron beam resist, *Opt. Lett.*, 18, 308–310, 1993.
27. Zhou Zhou and Sing H. Lee, Fabrication of an Improved Gray-scale Mask for Refractive Micro- and Meso-Optics, *Opt. Lett.*, 29 (5), 457, 2004.

28. Zhou Zhou and Sing H. Lee, *New Gray-Scale Technology Part (1)-Two-beam-current method for the e-beam writing and its application to high-resolution optical structures*, to be submitted to *Applied Optics* in late March, 2007.
29. R. Khalil et al., Optical clock distribution for high speed computers, *SPIE Proc.*, 991, 32–41, 1988.
30. J. Goodman, Optical interconnections for VLSI systems, *IEEE Proc.*, 72, 850–866, 1984.

Design of Illumination Systems

Introduction

Illumination optics is required in many varied system applications, including, for example, microscopes, projection systems, machine vision systems, industrial lighting. In optical systems where a light source is illuminating an object that has to be projected onto a screen as in a desktop projector, the design often requires a high brightness and uniformity across the image. High brightness implies a high collection efficiency of the light emitted from the source. Furthermore, these systems often require small packaging of the optical system.

Light sources have a wide range of types, shapes, and sizes, and the choice of the design of the illumination optics is very dependent on the source. Sources can be tungsten halogen lamps of different shapes, metal halide lamps, light-emitting diodes (LEDs), xenon lamps, frosted bulbs, different forms of arc lamps, or fusion (sulfur) lamps. Some sources have sufficient brightness and uniformity across their emitting area, and these can be imaged directly onto the object that has to be illuminated, but in most cases the sources need some kind of homogenization in the illumination optics to achieve the required brightness uniformity, while simultaneously minimizing throughput losses. The most common physical parameters that are used in photometry

to characterize the source and the illumination system are *flux*, *intensity*, *illuminance*, and *brightness*. *Photometry* deals with visual systems, and it is implicit that all relationships are weighted by the spectral sensitivity of the human eye. Clearly, if an object being viewed visually appears to be of a certain brightness, this must take into account the eye's spectral sensitivity. *Radiometry* is a closely related subject, which deals directly with power as it is emitted by a source, and ultimately irradiates a surface. The concepts and the basic theory are identical to those used in photometry; however, the units are totally different. We will define the basic photometric parameters and their corresponding units:

- *Flux* corresponds to power in radiometry, and it is the total power emitted by the source. The unit for flux is the lumen.
- *Intensity* is the flux per unit solid angle, or a solid angle flux density or the flux angular distribution, which assumes that the flux comes from a point source. The unit for intensity is the candela = lumen/steradian.
- *Illuminance* is the flux per unit area incident onto the surface being illuminated, or the area flux density. It does not relate to the angular flux distribution incident on the surface. The unit for illuminance is the foot-candela = lumen/ft², or lux = lumen/m².
- *Brightness (luminance)* is the flux per solid angle per unit area, or the area and solid angle flux density. The unit for brightness is the nit = candela/m².

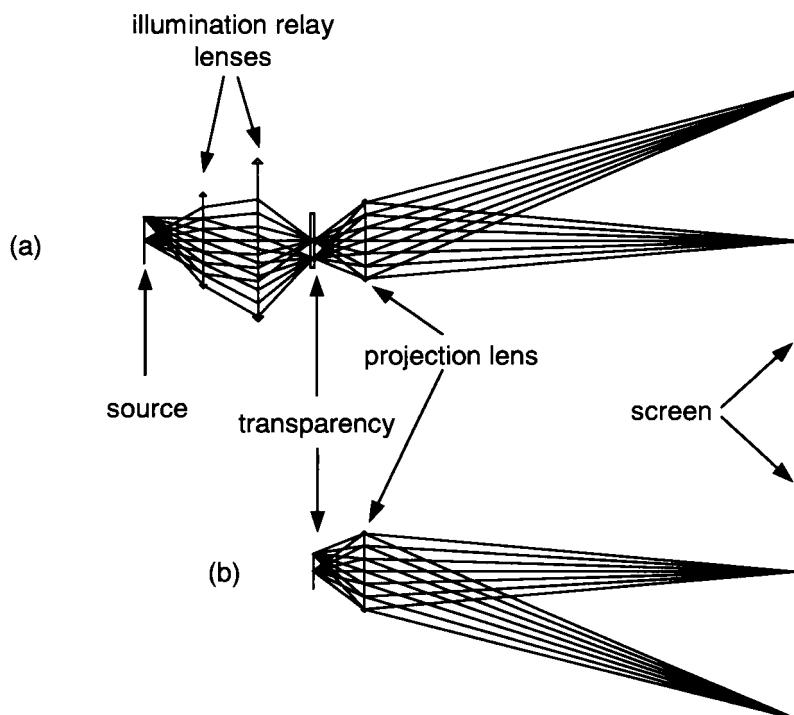
Köhler and Abbe Illumination

There are two classic approaches to the design of illumination systems. Most modern illumination system designs are modifications of one of these basic concepts.

Abbe illumination, which is sometimes called *critical illumination*, images the source directly onto the object to be illuminated, as shown in the paraxial model in Fig. 14.1. It is used in the cases when the source is sufficiently uniform for the system requirements. An illumination relay lens images the source onto the object. The object being illuminated can be film as in a movie projector, a 35-mm slide, or other similar transparency, or a transmissive or reflective LCD panel as in a desktop projector. Either



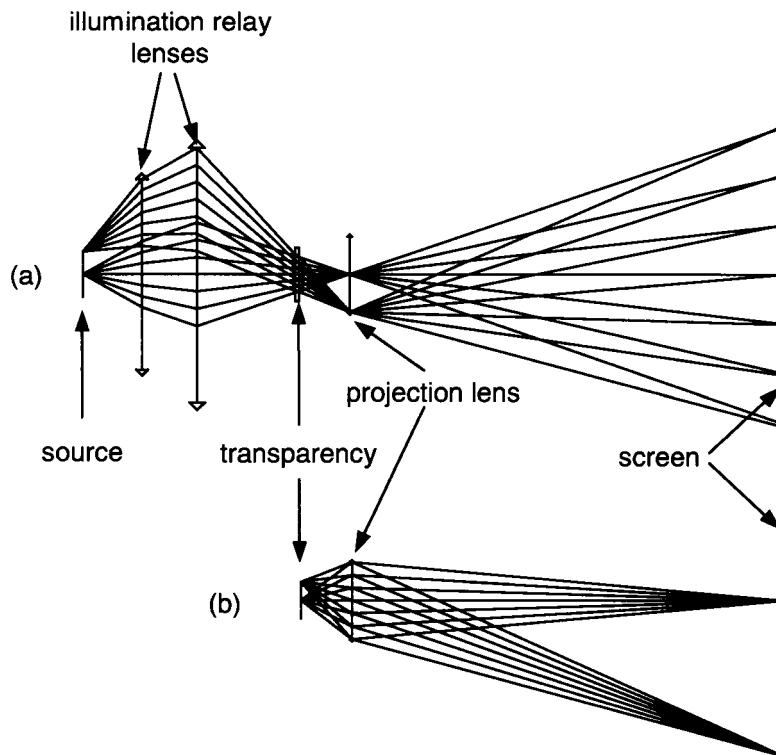
Figure 14.1
Abbe Illumination



the source must be inherently uniform as noted previously, or the condenser must have sufficient aberrations that blur the image of the source enough to eliminate the structure of the source at the transparency plane. The projection lens then images the object (transparency) onto the screen. At the same time, it images the image of the source onto the screen. This type of illumination works well with frosted bulbs or large sources, but it does not work with high brightness sources where high throughput is required and where there is significant structure in the source as with filament lamps and arc lamps. Generally, the resulting brightness and uniformity across the screen depends on the brightness and uniformity of the source, which means both spatial and the angular source uniformity.

In the case of a small bright, highly nonuniform source, such as an arc or a filament lamp, uniform brightness at the image is achieved with another form of illumination known as *Köhler illumination*, as shown in Fig. 14.2. Here, an illumination relay lens or condenser lens images the source into the *pupil* of the projection lens rather than onto the object being projected. This illumination relay system has its aperture stop at

Figure 14.2
Köhler Illumination



the location of the film, or the object, that is to be projected onto the screen, or into the eye. Since each point of the source illuminates the entire surface of the film, the film is by definition uniformly illuminated. Brightness nonuniformity with Köhler illumination can be caused by significant intense nonuniformity of the source.

Optical Invariant and Etendue

When a given light source emits a certain flux, not all of the emitted light reaches the screen or the detector. Some of that light is lost immediately after leaving the source. Since most sources emit into a large solid angle, often into a full sphere, we find that the first optical element, which may be a condenser system, has a difficult job of collecting and orienting or directing that light toward the aperture of the projection

optical system. Some of the light is lost in the optical system as absorbed, scattered, diffracted, or vignetted light. There may also be some components, such as the beamsplitters, filters, or polarizers, whose purpose in the system is to transmit only a certain type of light and block the rest.

There are many different terms that are related to the coupling of the light from the source to the screen. These terms are the *optical invariant*, *etendue*, *light-gathering power*, *throughput*, *angle to area*, and *area-solid angle product*.

Let us first take an example of an imaging system such as a telescope. The flux that is transmitted through a telescope goes through the aperture stop of the telescope shown in Fig. 14.3. If our telescope has an entrance pupil diameter D_{in} and the field of view is $2\omega_{\text{in}}$, the exit pupil diameter is D_{out} and the exiting field of view is $2\omega_{\text{out}}$, then

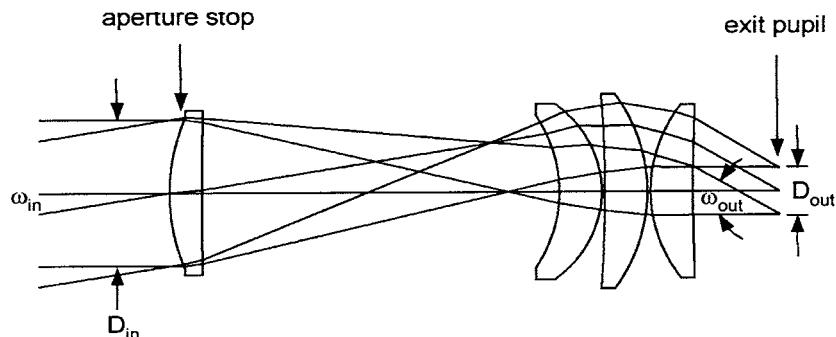
$$D_{\text{in}} \sin(\omega_{\text{in}}) = D_{\text{out}} \sin(\omega_{\text{out}})$$

In other words, optical invariant in the entrance pupil has the same value as in the exit pupil. This product maintains its value through the entire optical system. If we square the previous equation, we get

$$D_{\text{in}}^2 \sin^2(\omega_{\text{in}}) = D_{\text{out}}^2 \sin^2(\omega_{\text{out}})$$

The total flux that passes through the entrance pupil of the telescope will pass through the exit pupil if there are no losses within the system from absorption, vignetting, beamsplitting, filters, or scattering. This product $D_{\text{in}}^2 \sin^2(\omega_{\text{in}})$ is the measure of the throughput of the optical system and it is proportional to the area—solid angle product, or etendue. Etendue is, in effect, the conservation of flux within an optical system.

Figure 14.3
Flux at the Stop of a
Telescope and Tele-
scope Throughput



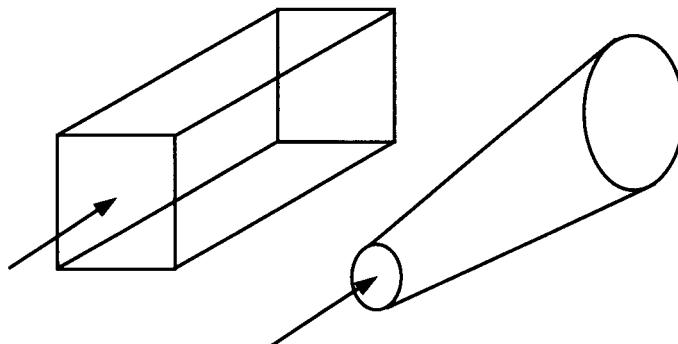
In a system with a light source, the goal is to use as much of the emitted light from the source as possible, and couple it into the optical system such as a projection lens for example. The etendue of the source (area of the source times the solid angle into which the source emits light) should be the same or slightly larger than the etendue of the projection lens. If the etendue of the source is significantly larger than the etendue of the projection lens, there are a lot of rays that are stopped by the apertures in the optical train, and these rays are lost and never reach the screen. On the other hand, if the projection lens is designed with the etendue larger than the etendue of the source, then the lens might be unnecessarily complex, with an underfilled aperture stop or underfilled field of view.

There are cases when the etendue is increased along the optical train. For example, if there is a need to polarize the light, then the unwanted polarization, instead of throwing it away, can be rotated at the expense of an increase in etendue. Another example is the presence of diffractive components in the system. If more than one diffractive order is used, the etendue in the system after the diffractive element is larger than before the diffractive element. Unfortunately, it is not possible to decrease the etendue without the light loss in an optical system.

The best correction of aberrations in most imaging optical systems is generally done at the center of the field of view. In illumination systems, on the contrary, the best aberration correction must be done at the edge of the field to get the sharp edge of the illuminated patch of light on the film, LCD or transparency.

Light pipes, as shown in Fig. 14.4, are commonly used as light homogenization components in illumination systems. Light pipes can be

Figure 14.4
Light Pipes



either hollow structures with reflective inside surfaces or they can be solid structures inside of which the light is totally internally reflected. Light pipes have two important roles in illumination systems. The first is that they are used as light homogenizers to change a spatial nonuniform distribution of the light at the input of the pipe into a uniform output. By definition, the etendue at the input surface of the light pipe is equal to the etendue at the output of the pipe.

The second very important function of light pipes is that tapered light pipes can change the angle of the input cone of light into a cone that can be accepted by the system into which the light from the source is injected. The optical invariant at the input of the pipe is equal to the invariant at the output

$$\frac{D_{\text{out}}}{D_{\text{in}}} = \frac{\sin(\omega)_{\text{in}}}{\sin(\omega)_{\text{out}}}$$

This formula gives the angle conversion in the case of a tapered light pipe and this is shown in Fig. 14.5. A tapered light pipe with multiple inside reflections is exactly like a kaleidoscope! Changing the angle between the rays and the optical axis after each internal reflection, it creates an array of virtual sources. This helps to spatially homogenize the brightness at the output of the pipe. Some shapes of the input (output) pipe surfaces spatially homogenize the light better than the others. Shapes such as a rectangle, a triangle, or a hexagon are the tileable shapes after unfolding due to internal reflections, and

Figure 14.5
Angle Conversion in
a Tapered Light Pipe

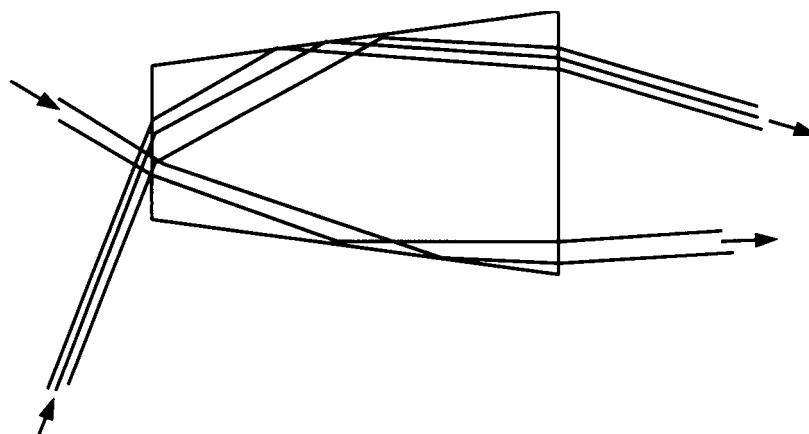
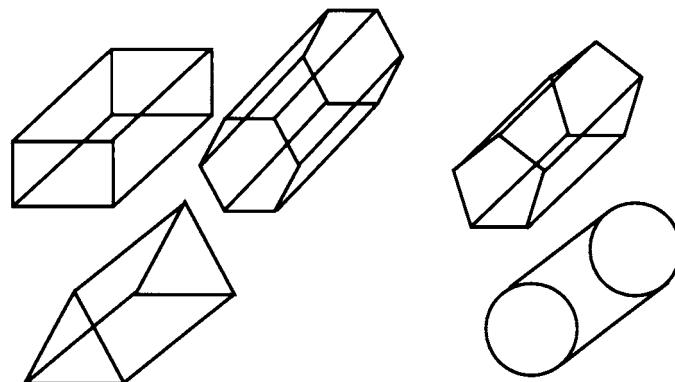


Figure 14.6
Different Shapes of
Light Pipes

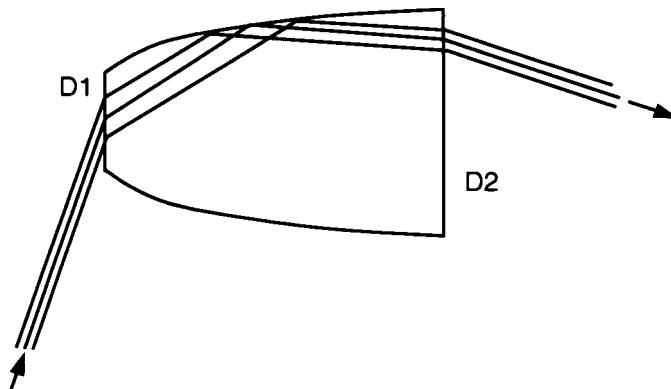


shapes that are good
spatial homogenizers of
light at the output

shapes that are bad
spatial homogenizers of
light at the output

they spatially homogenize the light very well. Different shapes are shown in Fig. 14.6. Straight walls of these pipes are good enough to give a uniform output. Other types that are not tileable, such as a circular input surface, do not homogenize the light as well, even when the pipe is very long. In the case of a circular input, improvement in uniformity is achieved with a pipe wall that is parabolic rather than a cone. This type of pipe is called a *compound parabolic concentrator* (CPC), as shown in Fig. 14.7.

Figure 14.7
Compound Parabolic
Concentrator (CPC)



Other Types of Illumination Systems

The most efficient way to collect the light from a small source is to place the source at the focus of a parabolic or elliptical reflector. These reflectors collect the light emitted by the source into a large solid angle, and either collimate the light in the case of a parabolic reflector, or focus it to the second focal point of the ellipse shown in Fig. 14.8. There are different ways to reduce the nonuniformity of the image of the source and get a smooth illumination. One way of smoothing in collimating reflectors, such as flash lamps or street or automobile lighting, is the wedged reflector, where the reflector has a basic parabolic form, but it is divided into a lot of wedge-shaped segments. The other way is to combine the collimated light from the reflector with a smoothing plastic element located in front of the reflector, consisting of the extruded array of prisms or structures with combined prismatic and sinusoidal profiles.

In the case of a film projector, the goal is to send the light from the source into the rectangular area of the film, with the numerical aperture of the illumination matched to the numerical aperture of the projection lens. There are two common ways of getting a rectangular uniformly illuminated area in the film plane from a highly nonuniform, nonrectangular source.

The first method is to initially collect the light from the source with a reflector, then focus it down onto the input surface of the rectangular light pipe, as in Fig. 14.9. The magnification of the tapered light pipe is

Figure 14.8
Parabolic and
Elliptical Reflectors

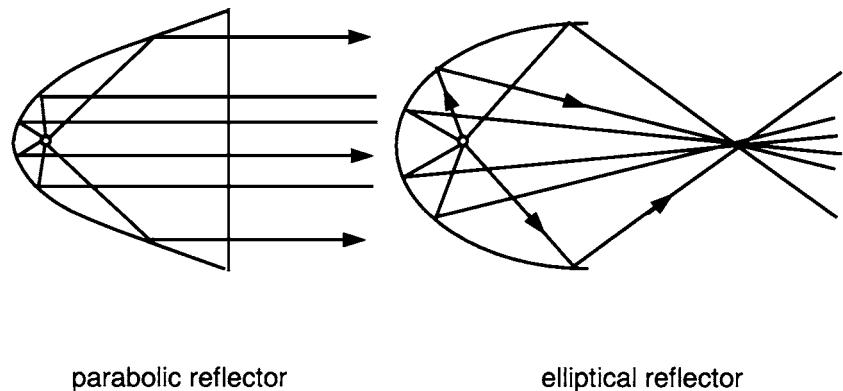
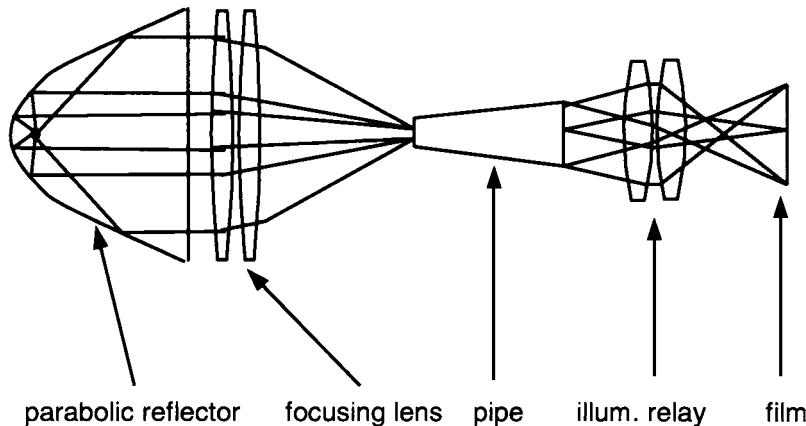


Figure 14.9
Illumination System
with a Tapered Light Pipe



chosen such that the etendue at the output from the pipe is equal to the etendue of the optical system that relays the rectangular pipe output surface onto the plane of the film. The pipe has to be sufficiently long to get the spatially uniform output. A good rule of thumb is that the output will be uniform if there are at least three ray reflections along the pipe. The larger the magnification of the pipe, the smaller the minimum length of the pipe needed to achieve the uniform output.

The second method is often used in the illumination systems of front desktop projectors with transmissive image panels and with projection lenses of $f/3$ and above. It uses lenslet arrays along with a polarization recapture plate. This is shown in Fig. 14.10. The light from the source is collimated with the parabolic reflector. The expanded and collimated beam then goes through the lenslet array (lenslet array is shown in Fig. 14.11). Each lenslet focuses the beam inside the lenslet of the second lenslet array producing an image of the source. The light that illuminates the image panels has to be linearly polarized. After the second array, there is an array of polarization beamsplitters, which splits the light into two orthogonal linearly polarized beams. One of the beams then undergoes the rotation of its plane of polarization, and two beams emerge in the same polarization state out of the polarization recovery array. The second lenslet array, together with the focusing lens, images the lenslets (which are rectangular) of the first array onto the image panel. The focusing lens superimposes images of the lenslets of the first array in the plane of the image panel. In summary, a circular distribution of flux from the parabola is sampled by rectangular elements of the lens arrays and is then superimposed at the

Figure 14.10
Illumination System
with Lenslet Arrays
(Only the Rays Passing
through One
Lenslet Are Shown)

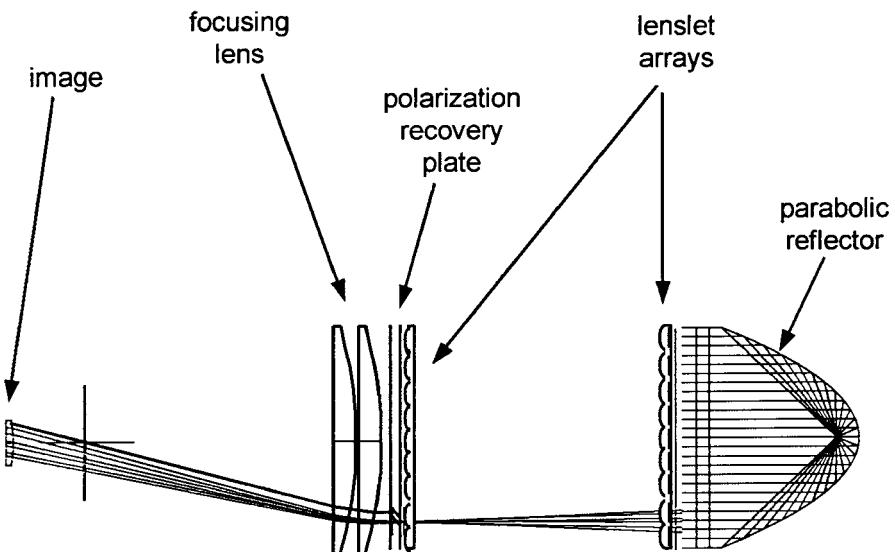
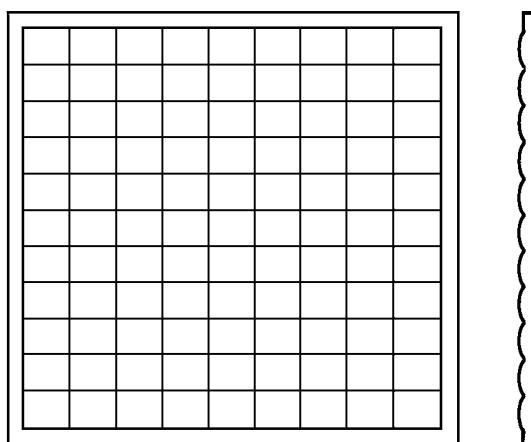


image panel, thus homogenizing the nonuniform parabolic output and transforming the geometry to match that of the image panel. An illumination system with a light pipe has a high throughput and small packaging. However, mounting of the light pipe is a problem, since all pipe surfaces are the optical surfaces, and any contact with the sides of the pipe frustrates the total internal reflection, resulting in a loss of light. Although the system with lenslet arrays requires more space, it accommodates the polarization recovery and increases the throughput 30 to 40%.

Figure 14.11
Lenslet Arrays



This page intentionally left blank

Performance Evaluation and Optical Testing

The quantitative characterization of optical performance, or image quality, is extremely important. Generally, the optical design engineer plays a key role in system testing, and for this reason we feel it is important to include the basics of optical testing in this book. Testing can range from the somewhat simplistic bar target to the more sophisticated means for characterizing the modulation transfer function (MTF).

Testing with the Standard 1951 U.S. Air Force Target

The simplest form of resolution target is perhaps the white picket fence shown in Fig. 15.1. The image of the fence consists of alternating bright and dark bars as formed by the white pickets and the dark background between the pickets. If we image this fence with a camera lens, the image will be demagnified by approximately the ratio of the camera

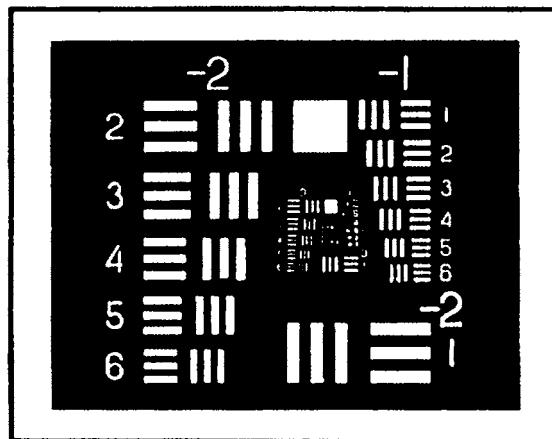
Figure 15.1
The White Picket
Fence Analogy to a
Bar Target for Optical
Testing



focal length to the distance to the fence. Let's assume the fence pickets are 75-mm wide and that we are imaging the fence with a 50-mm focal length lens from a distance of 20 m. The magnification is therefore $50/20,000 = 0.0025\times$. The fence is 150 mm/picket pair, or equivalently 150 mm/line pair. This equates to 0.006667 line pairs/mm. At the image formed by our lens, this becomes 0.375 mm/line pair, or 2.667 line pair/mm. Most camera lenses will resolve this spatial frequency just fine as it is a rather low spatial frequency. Let's now move our lens to a distance of 200 m. Here the magnification is 0.00025 \times and the spatial frequency becomes 26.67 line pair/mm. A reasonably good camera lens will have a contrast of approximately 50% or higher at this spatial frequency. Needless to say, as the lens moves further from the fence the spatial frequency, in line pairs per millimeter increases and the contrast decreases as a result of aberrations, diffraction, assembly and alignment errors, and other factors.

If we had no other metric, a white picket fence would be a reasonable target to use for lens performance testing and characterization. In the laboratory, the most basic means for measuring image quality is through the use of the so-called 1951 U.S. Air Force Target. This form of target is readily available, low in cost, and easy to use. A typical Air Force target is

Figure 15.2
The Standard 1951
U.S. Air Force Target



shown in Fig. 15.2. The legend for the target is shown in Table 15.1 where it is evident that the target is divided into groups and elements so, for example, group 2, element 4 is a bar pattern of 5.66 line pair/mm.

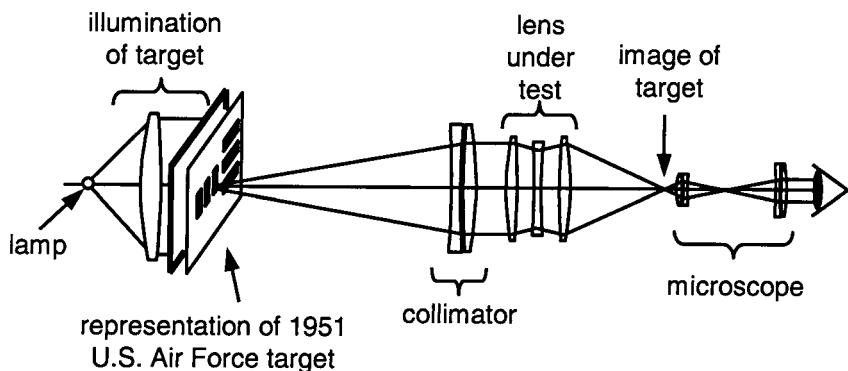
How do we use an Air Force type target? Let us consider the example of measuring the performance of a 100-mm focal length 35-mm camera lens. Let us further assume that we must “resolve” 50 line pair/mm. As will be discussed in Chap. 22, this is a reasonable value for such a lens. We now construct a test setup shown in Fig. 15.3. We locate an Air Force target at the focus of a collimator lens. The collimator is used to simulate an infinite object distance. It is *critical* that the quality of the collimator *must* be independently validated and demonstrated to be better than the level of performance we are looking for. Generally, the focal length

TABLE 15.1

Legend for 1951
Standard U.S. Air
Force Resolution
Target, lp/mm

Element Number	Group Number									
	-2	-1	0	1	2	3	4	5	6	7
1	0.25	0.5	1.0	2.0	4.0	8.0	16.0	32.0	64.0	128.0
2	0.28	0.561	1.12	2.24	4.49	8.98	17.95	36.0	71.8	144.0
3	0.315	0.63	1.26	2.52	5.04	10.1	20.16	40.3	80.6	161.0
4	0.353	0.707	1.41	2.83	5.66	11.3	22.62	45.3	90.5	181.0
5	0.397	0.793	1.59	3.17	6.35	12.7	25.39	50.8	102.0	203.0
6	0.445	0.891	1.78	3.56	7.13	14.3	28.51	57.0	114.0	228.0

Figure 15.3
Using the Air Force Target to Test a Camera Lens



of the collimator lens should be at least a few times longer than the focal length of the lens under test (a factor of 3 is the minimum). The collimated light now enters the lens under test and the image of the target comes to focus in the image plane as shown.

In order to compute which pattern corresponds to 50 line pair/mm, we can simply multiply 50 by the magnification from the target to the final image. This magnification is the focal length of the camera lens divided by the focal length of the collimator. Assume we have a collimator with a 1-m focal length. This gives us a magnification of $0.1\times$, which means that 5 line pair/mm is the spatial frequency on the target corresponding to 50 line pair/mm at the focus of our camera lens. Group 2, element 3 is 5.04 line pair/mm and is a sufficiently close match.

We now magnify this image with a microscope of suitable magnification so that the 50 line pair/mm is sufficiently well magnified so as not to be limited by the resolution of the eye. In order to compute the required magnification, first we need to find what angular subtense our 50 line pair/mm equates to: $50 \text{ line pair/mm} = 0.02 \text{ mm/line pair}$, which when viewed at a nominal distance of 254 mm corresponds to $0.02/254 = 0.00008 \text{ rad}$. The eye resolves approximately 1 min of arc, which is 0.0003 rad . Our target is thus about 4 times smaller than what the eye can resolve, and this is not acceptable. If we magnify the target by $40\times$ with the microscope, it will be seen at 10 times the resolution limit of the eye, which should be just fine. Ten arc minutes equates to 1 cm at a distance of 3 m. Thus, a $40\times$ microscope should be used. To determine the magnification of the microscope, simply multiply the magnification of the objective lens by the magnification of the eyepiece. Thus, a $4\times$ objective and a $10\times$ eyepiece will do the job for us. We may want to use a higher

magnification to be sure that the eye is in no way limiting and for more comfortable viewing of the image. The microscope objective has to have a numerical aperture that is larger than the numerical aperture of the lens under test.

The person performing the test now visually views the bar pattern through the microscope and judges whether he or she can distinguish three separate bars in the pattern in group 2, element 3 (for both horizontal and vertical orientations). If the answer is “yes,” the lens passes, if the answer is “no,” the lens fails. This test is indeed quite simple and easy to set up. The disadvantage of this form of test is that a real person can only judge whether the bar pattern can or cannot be resolved, not the level of contrast or sharpness of the image. Also, the test is somewhat subjective and sometimes different people will get different answers. A common rule of thumb is that the eye can resolve a modulation of approximately 5%, or 0.05. This is a useful metric, but what about the lens’ ability to produce a good image with a higher contrast, say at 20 line pair/mm? The lens will likely resolve this spatial frequency fine; however, the user cannot judge the level of modulation or contrast. The only thing the user can judge is whether a specific spatial frequency is or is not resolvable or distinguishable into three bars.

The Modulation Transfer Function

The modulation transfer function provides a more quantitative measure of the image quality of an imaging lens system than a bar target. We discussed the basics of MTF in Chap. 10. From a laboratory standpoint, there are two principal ways of measuring MTF. First, we can use sinusoidal patterns of different spatial frequencies and image them with the lens under test onto a CCD sensor to characterize the resulting modulation as a function of spatial frequency. If you know the modulation of your target pattern, you can compute directly the modulation transfer function.

A more robust and easier way to measure MTF is by computing the Fourier transform of the line-spread function, which gives the MTF directly. The Fourier transform is a widely used mathematical transformation whereby we can compute the transformation from the spatial domain to the frequency domain, thereby giving the MTF directly. A lab setup such as in Fig. 15.4 is used, and is explained here. The test setup shown is for a finite conjugate lens, which is $f/20$ on its object side and $f/2.8$ on its image side:

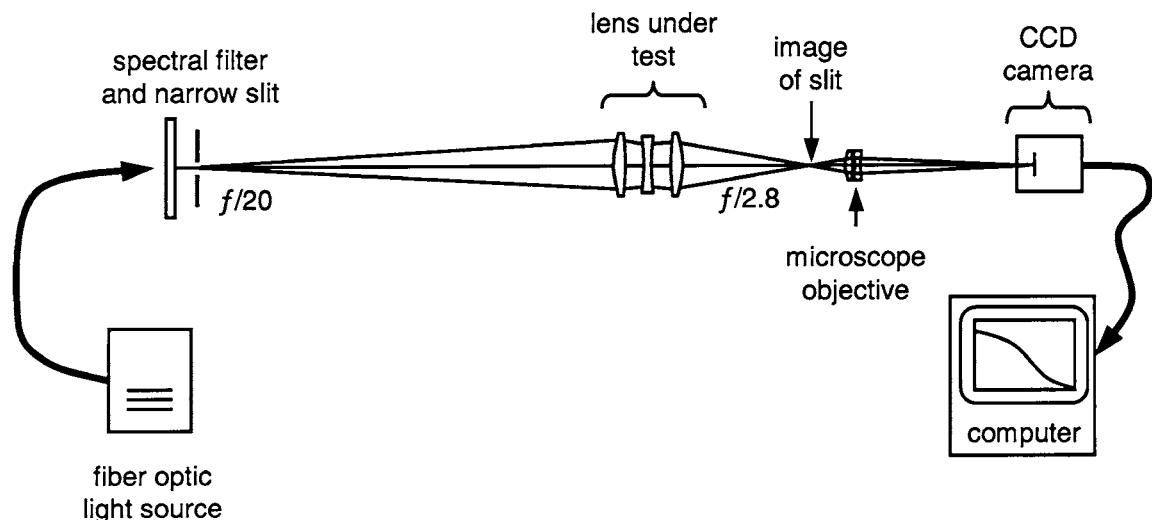


Figure 15.4
Example of Lab Setup to Measure MTF

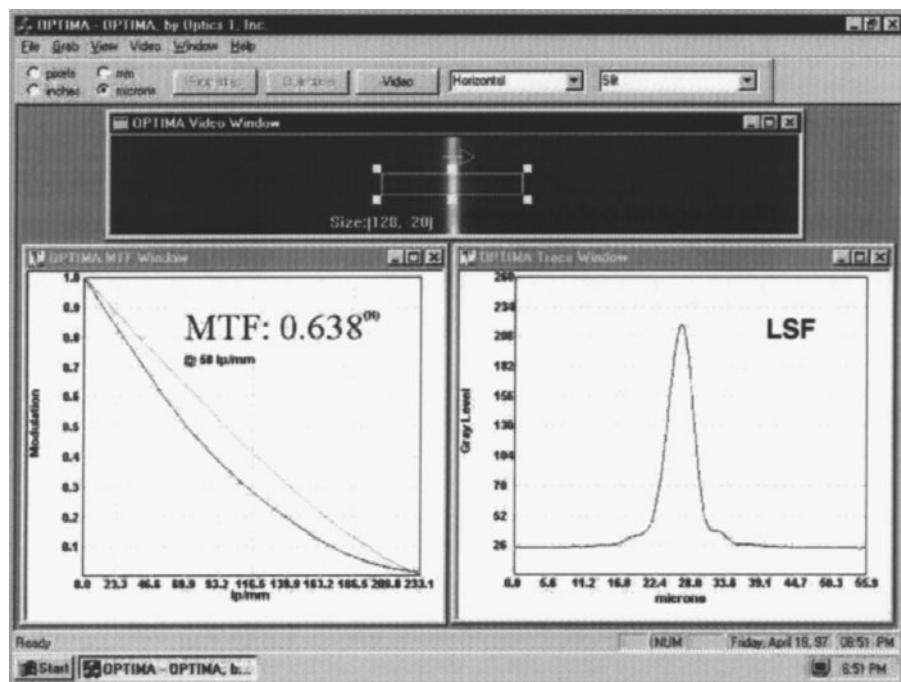
- A light source with an appropriate spectral filter illuminates a narrow slit. The slit width should be narrow enough so that its finite width does not affect the results. As we know, the diameter of the Airy disk diffraction pattern is approximately equal to the f/number , in micrometers. A good rule of thumb is that if our slit width is 25% or less of this value, then its effect on the resulting Fourier transform will be negligible. For our $f/20$ object cone, the slit width required is approximately $5 \mu\text{m}$ or less since the Airy disk diameter would be about $20 \mu\text{m}$. If we cannot find a sufficiently narrow slit, the software can divide the resulting MTF by the Fourier transform of the rectangular slit, which is the MTF of the slit itself. This, in effect, divides out the effect of the finite slit width.
- The lens under test is now positioned on the optical bench.
- The image of the slit is now magnified with a high-quality microscope. The profile of the intensity distribution in the image of the slit, in the direction normal to the slit, is called the *line-spread function* (LSF). The magnified image of the slit is imaged onto a CCD array.

- The analog or digital output of the camera is input into a computer with a frame grabber, where the real-time video image of the slit is displayed in one of the windows, as shown in Fig. 15.5. A scan through the gray levels (line-spread function) is displayed in another window.
- The Fourier transform of the LSF is computed and is displayed in the remaining window. The perfect system MTF or diffraction-limited MTF is shown for reference and is computed based on user input of the central wavelength and the f/number .

The MTF is one of the most useful means for characterizing the optical performance of an imaging system. Most scenes consist of objects of many spatial frequencies, and the MTF data tell us how the modulation of the object is transferred from the object to the image as a function of the many varied spatial frequencies in the scene.

All lens design computer programs allow for the computation of the MTF for a given design. It is important to remember that these results

Figure 15.5
Example of MTF
Measurement
Software

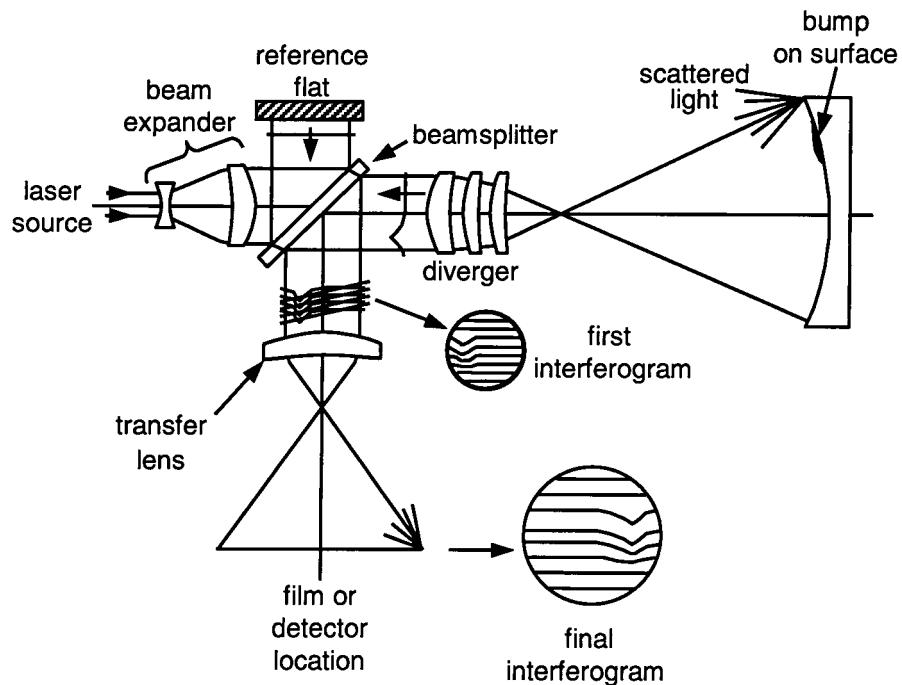


are for the design prescription, or paper design, and do not include the effects of fabrication, assembly, and alignment errors. This nominal design thus needs to have sufficiently higher MTF to account for the MTF degradations due to these manufacturing errors.

Interferometry

Interferometry allows us to measure quantitatively the optical path difference of a lens or alternatively, the departure of a curved or flat surface from its nominal shape. While the geometry of many interferometers is different, their basic functionality is similar in that the wavefront from a perfect reference surface, such as an optical flat, is compared and interferes with the wavefront from the lens or surface under test. To illustrate just how an interferometer works, consider Fig. 15.6, where a Twyman-Green interferometer is shown:

Figure 15.6
Twyman-Green
Interferometer



- Light from a monochromatic and coherent light source, such as a HeNe laser, is input into a beam expander. The beam expander expands the approximately 0.8-mm-diameter beam to a diameter in the order of 25 to 50 mm. While the beam expander should be of high quality, it is not necessary that it be *perfect* as any small errors will cancel since the beam goes to both the reference arm as well as the test arm of the interferometer, and small wavefront errors will cancel each other.
- The expanded laser beam is incident onto a nominally 50–50 beamsplitter plate. Let us assume that the first (left-hand) surface has the beamsplitter coating and the opposite side is antireflection coated.
- Fifty percent of the light now travels upward to the reference flat. This mirror must be near perfect and is often in the order of $\lambda/40$ wave *P-V*. The light reflects downward from the reference flat and 50% of it passes through the beamsplitter, with the rest going back toward the laser. This is the so-called reference arm of the interferometer.
- The remaining 50% of the light, which passes through the beamsplitter, enters what is known as a *diverger* or *transmission sphere*. This is the test arm of the interferometer. We will assume that we are testing here a spherical mirror. The diverger is, in effect, a perfect lens that creates a perfect Airy disk at its focus and creates perfectly diverging wavefronts following the intermediate image. The purpose of the diverger is to create a wavefront that perfectly matches and nests into the nominal shape of the mirror under test, which means it appears that the wavefront emerges from the point coincident with the center of curvature of the mirror under test. If the mirror were a paraboloid or some other nonspherical shape, then a *null lens* is used instead of a diverger, with the same goal of creating a wavefront which nests into the nominal surface under test. We will show the use of a null lens in Chap. 25.
- We will assume that the mirror under test has a small bump on it, as shown, but is otherwise perfect. A portion of the wavefront will first hit the top of this bump and reverse its direction. The rest of the wavefront will now travel to the mirror surface and then turn around. Meanwhile the part of the wavefront that hits the top of the bump is already heading back toward the diverger. The bump

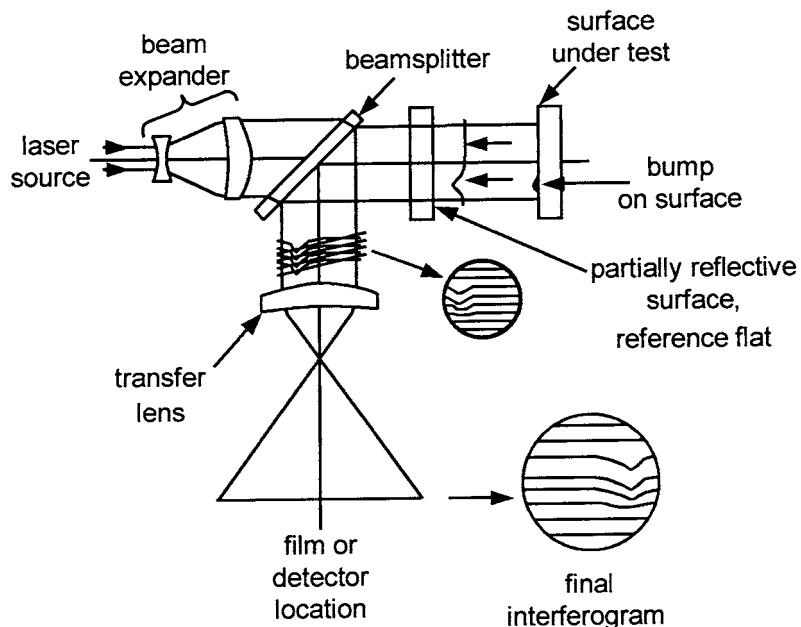
in the wavefront will thus be a factor of 2 times as large as the physical height of the bump itself.

- The return wavefront passes through the diverger and is once again a series of plane parallel wavefronts, with the exception of the bump.
- Fifty percent of the light incident onto the beamsplitter will now reflect downward, with the rest going toward the laser. At this point, we have two wavefronts that are located between the beamsplitter and the transfer lens, one from the reference flat and the other from the mirror under test. Since the light is monochromatic and coherent, the light will interfere. What this means is that for regions where the two wavefronts are in phase we see a bright fringe and for regions where they are 180° out of phase we see a dark fringe. Since the wavefront from the reference arm of the interferometer is essentially perfect, any deviations are from the test arm, specifically from the mirror under test. If we were to place a white card in the beam here, we would see interference fringes indicative of the departure of the mirror under test from perfect sphericity. There is only one problem—due to scattering from the edge of the mirror under test we may see artificially curved fringes at the outer periphery of the interferogram. This effect could be interpreted as a turned up or down edge on the mirror, which is false.
- In order to eliminate the false turned up or down edge phenomena described, a *transfer lens* is used. This lens, or lens group, serves a dual role. It images the surface under test onto the sensor (CCD film, etc.), thus bringing the scattered light back to the edge of the interferogram where it belongs to eliminate false fringes. It also matches the size of the image to the sensor. If this is done properly, the final interferogram will show crisp edges with no artificial edge effects. Figure 15.6 shows how the scattered light at the edge of the surface under test is reimaged onto the interferogram plane by the transfer lens.

Another common form of interferometer is the Fizeau interferometer, as shown in Fig. 15.7. This system is similar to the Twyman-Green interferometer; only the reference beam is created from the partial reflection from a reference flat, which is also used in transmission.

It is important in any of the interferometer tests that the two beams (the reference and the test beams) are of approximately the same intensity in order to maximize the fringe contrast.

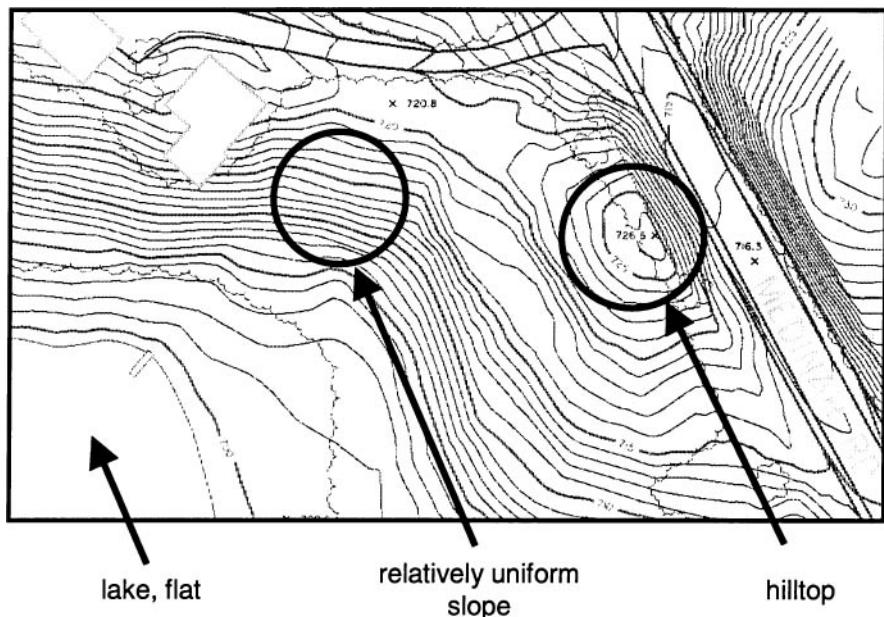
Figure 15.7
Fizeau Interferometer



The interpretation of an interferogram is very much like interpreting a topographical map with contours of equal altitude or elevation. In an interferogram, the fringes are lines or contours of equal height just as with a topographical map. Consider the topographical map shown in Fig. 15.8. The lake at the lower left is flat or level; hence there are no contour lines. The top center circle contains nearly straight contours, which indicates a flat region which is also tilted or sloped, and the circle at the right of the hilltop shows a somewhat dome-shaped hilltop which has a rapid falloff to the right. The interpretation of an interferogram is exactly the same as that of a topographical map with the only difference being the scale of the contours. In a topographical map the contour levels may be in units of meters or feet, and in an interferogram the units are in wavelengths of light.

Now consider Fig. 15.9 where we show four interferograms of a nominally flat mirror. Each fringe is due to one-half wave of surface departure from flatness, which results in one wave to the reflected wavefront, hence one fringe. The upper left interferogram shows a surface, which for the most part is tilted left to right. The upper right interferogram is quite flat, with a residual tilt from an 11 o'clock to a 5 o'clock direction.

Figure 15.8
Topographical Map
Analogy to Interferogram



And the lower left interferogram shows a saddle-shaped residual. It is important to note that interferograms contain no sign information whatsoever with respect to what is high and what is low on the surface. This information *must* be obtained when the part is being tested, as it is impossible to derive the sign information at a later time. The saddle could be down in the 10 o'clock and 4 o'clock directions and up in the 1 o'clock and 7 o'clock directions. Alternatively, the up and down directions could be reversed. In fact, it is physically possible (although unlikely) that the surface only departs up (or down) in the four regions.

Other Tests

There are many other optical system testing methods, and the more important of these will be briefly outlined here:

- The *star test* is where you view the image of a point object (a pinhole) similar to a star through a microscope visually (or via a video camera). If a reasonably narrow-band filter is used, the image should be that



Figure 15.9
Typical Interferograms of Nominally Flat Mirrors

interferograms of thin aluminum mirror
(both interferograms are of same mirror)



interferogram of thin aluminum mirror with astigmatism or saddle



interferogram of thin beryllium mirror

of an Airy disk if the system is diffraction limited. This is an extremely sensitive test since very small asymmetries in the Airy disk are quite evident. The human eye is sensitive to many orders of magnitude in dynamic range, which makes the test quite robust. If the microscope is now focused inside and then outside of best focus, you will see a disk with concentric rings which becomes larger as you move further from best focus. This ringed disk should appear similar both inside and outside of best focus. If it is not, then you have a residual of spherical aberration. While not quantitative, the star test is an extremely sensitive test.

- The *Hartman test* is often used at observatories to test telescopes. A mask with a grid of small apertures is placed in front of the primary mirror, and the resulting imagery is recorded at several

through-focus locations. This is the analogy of creating a series of through-focus spot diagrams, and the residual aberrations can be derived using appropriate software.

- Other forms of optical system testing relate more to alignment of the system as a whole as well as its subcomponents or groups. Measurement of focal length, distortion, and other lens metrics is also important. We often use alignment telescopes, laser beams, and other methods for assuring that our optics is sufficiently well aligned.

Tolerancing and Producibility

Introduction

Most of the material presented in the earlier chapters of this book is associated with achieving the optimum lens design. There has been a lot said over the years about lens design optimization theory and algorithms, global search algorithms, aberration theory, and other related topics. These are all directed toward achieving the optimum lens performance for what we sometimes call the “paper design.” The performance of the paper design is that of the design prescription, or the theoretical design, with the effects of absolutely no manufacturing errors of any kind included. While the performance of the paper design is indeed important, the effects of real-world hardware-related manufacturing errors and tolerances can and will, by definition, alter and degrade the level of performance from the theoretical paper design performance. We often find that the degradations due to manufacturing errors can many times even surpass the image degradations of the nominal design itself.

Tolerancing is the science (and art, to some extent) of distributing and error budgeting the manufacturing tolerances of all optical and optomechanical components and dimensions throughout the system to assure that your system will meet its required level of optical performance at a reasonable cost.

Unfortunately, there is little, if any, direct correlation between the performance of the paper design and the robustness or insensitivity of the

design with respect to the level of manufacturing errors or tolerances. For example, we may have a design where steep bendings and high angles of incidence has allowed for the effective balancing of higher orders of aberration, thus yielding a high level of performance for the paper design. Unfortunately, this design may be extremely sensitive to tolerances due to the higher angles of incidence and the presence of higher-order aberrations. On the other hand, a different design configuration for the same lens requirements may have significantly reduced angles of incidence, reduced higher-order aberrations, and may result in a somewhat lower level of performance for the paper design. However, due to the reduced angles of incidence, this design may be less sensitive to manufacturing errors and tolerances. An example of this was shown in Chap. 9 where we discussed several designs submitted to the 1980 International Lens Design Conference. One of these designs was a very compact lens with small angles of incidence at the surfaces, while one of the other designs had extremely large angles of incidence on several surfaces. This latter design will, in all likelihood, have tighter manufacturing tolerances.

The key point is that manufacturing errors in the form of fabrication, assembly, and alignment errors can be extremely important and are often a major contributor to the overall level of performance of an optical system, even if the paper design is excellent.

In tolerancing an optical system, we need to assign tolerances to all optical and mechanical components within the system. This includes all lenses and/or mirrors as well as the mechanical components, which directly or indirectly support the optics. The overall goal for the system is that the optical performance is met (MTF and/or other image quality criteria), optical component costs are minimized, assembly and alignment costs are minimized, and yields are maximized. Tolerancing is necessary whether you are producing a lens in a high-production environment or a one-of-a-kind lens.

Ultimately, the goal of the tolerancing effort is to aid in establishing a performance error budget whereby you can, with confidence, predict the expected level of optical performance.

What Are Testplates and Why Are They Important?

Prior to embarking on an extensive tolerance analysis, your design needs to be completed and finalized. One of the very last steps in this process is

that of matching radii to existing testplates or tooling. Virtually all optical shops have in their inventory hundreds of so-called testplates. These testplates (sometimes called *test glasses*) are a convex and concave mating pair of tooling with radii ranging from very short to very long. They are often made of low-expansion Pyrex glass, they have very low surface irregularity, and their radii have been measured to a high level of precision.

Once a design is finalized and the shop to manufacture the optics has been selected, the designer should proceed to match as many radii as possible to existing testplates. Let's take an example: Assume that we have a six-element double gauss lens with 10 different radii. Further, assume that after we compared the radii in the final design to the testplate list of the selected lens vendor, we found that the closest radius was 25.21 mm, and our vendor has a testplate of radius 25.235 mm, a difference of only 2.5 μm . Five sections of the testplate list from OPTIMAX is shown here to give you an indication of the number of plates available and the density, or closeness, of radii in the different radii regions.

Plate ID	Radius	Diameter	CC	CX
1	1.5000	2.5000	X	X
2	2.0000	3.2000	X	X
3	2.0000	3.9000	X	X
4	2.0470	3.5000	X	X
5	2.0470	4.0000	X	X
6	2.5150	3.2000	X	X
7	2.5150	5.0000	X	X
168	10.0150	14.2000	X	X
169	10.0150	14.9000	X	X
170	10.0600	13.2000	X	X
171	10.0600	14.9000	X	X
172	10.1800	13.9000	X	X
173	10.1800	14.7000	X	X
174	10.2240	13.5000	X	X
175	10.2240	16.6000	X	X
176	10.3100	17.0000	X	X

Plate ID	Radius	Diameter	CC	CX
177	10.3100	15.0000	X	X
178	10.3900	13.6000	X	X
179	10.3900	14.9000	X	X
180	10.5000	13.4000	X	X
181	10.5000	13.7000	X	X
483	25.0550	40.0000	X	X
484	25.0550	41.1000	X	X
485	25.1000	17.8000	X	X
486	25.1000	16.8000	X	X
487	25.2350	38.3000	X	X
488	25.2350	44.3000	X	X
489	25.2600	33.5000	X	X
490	25.2600	44.8000	X	X
491	25.3800	38.7000	X	X
492	25.3800	41.7000	X	X
493	25.4050	37.7000	X	X
494	25.4050	41.8000	X	X
495	25.4460	33.1000	X	X
496	25.4460	35.0000	X	X
948	75.3200	57.6000	X	X
949	75.3200	61.2000	X	X
950	75.5000	56.2000	X	X
951	75.5000	44.7000	X	X
952	75.8260	70.0000	X	X
953	75.8260	71.5000	X	X
954	76.2950	63.0000	X	X
955	76.2950	63.0000	X	X
956	76.4350	62.0000	X	X

Plate ID	Radius	Diameter	CC	CX
957	76.4350	63.7000	X	X
958	76.6150	49.9000	X	X
959	76.6150	52.0000	X	X
960	77.0930	69.0000	X	X
1491	500.2100	85.0000	X	X
1492	500.2100	87.0000	X	X
1493	501.5200	78.3000	X	X
1494	501.5200	56.6000	X	X
1495	508.1400	87.3000	X	X
1496	511.8840	88.0000	X	X
1497	511.8840	88.0000	X	X
1498	514.5900	69.3000	X	X
1499	514.5900	69.8000	X	X
1500	520.7050	68.3000	X	X
1501	520.7050	68.9000	X	X
1502	528.5250	102.4000	X	X
1503	528.5250	103.5000	X	X
1504	537.4110	73.6000	X	X
1505	537.4110	75.4000	X	X

We then proceed to change the radius in the design to the testplate value of 25.235 mm and then freeze it from any further changes. In effect, we will reoptimize the lens while constraining this radius to exactly match the testplate. All other variables and constraints of the design are the same as for our final optimization cycles, and our error function remains the same. After the first radius has been matched, we once again search for the closest radius of those remaining to any on the testplate list and match it to this testplate radius. Note that during the reoptimization process following the first testplate insertion, all of the remaining radii will change or “shuffle” a small amount, giving a whole new scenario with respect to which surface is now closest to an existing testplate. This process continues until we have matched all of the radii to existing testplate radii.

It is our experience that in most cases 100% of all radii should be able to be matched to existing testplates. If for some reason you have problems with the last one, two, or three radii, you can release several radii that you have already matched and then match the problem radius or radii, with the intent of matching 100% of all radii to existing testplates. A very important point to keep in mind is that when we are matching our last one or two testplates, there are barely enough variables to correct the aberrations, much less any constraints such as focal length. Thus, for the last few testplates we highly recommend either to allow some or all of the element thicknesses to vary (in addition to the already varying air spaces), or alternatively we may need to release or relax one or more of the system constraints such as focal length. Any further changes in the design will likely cause a negligible change in focal length. We had a system some years ago where we inadvertently left the constraint on focal length during the match of the final testplate. In order to meet this constraint, the design took on a highly degraded level of optical performance, which would have caused it to fail its performance specification. During the testplate fit, you may elect to match the closest or the farthest radius first; this is your choice. Our experience shows the closest algorithm to work just fine.

Most software programs have an automated testplate fitting routine, and these are quite robust and work well. But use them with care, as they are quite automated and you could run into one of the problems cited previously if you are not careful (such as the match of the last one or two testplates, producing poor performance due to overconstraining a first-order parameter such as focal length). If you do have a particular problem with matching a specific radius, you might consider running your optimization and varying this surface radius along with a constraint to match it to the testplate. By working the match into the optimization, your chances are best.

Why do we emphasize so strongly matching all radii to existing vendors, testplates, and how does this relate to tolerancing? The reasons are several:

- Testplates cost several hundred dollars or more each, and they take time to manufacture, perhaps up to several weeks.
- Testplate radii can be measured more accurately than they can be manufactured. For example, a radius of 25 mm can be *manufactured* to within about ± 0.025 mm (precision level for manufacturing); however, it can be *measured* to within about ± 0.00625 mm or less. We all realize that, given enough time and money, we could of course manufacture the radius to within the ± 0.00625 -mm tolerance or even better; however, this is generally not

economically feasible. Thus, matching testplates is like an insurance policy whereby your level of confidence in your lens working as predicted is enhanced after matching the radii to existing testplates.

When you perform your final tolerance analysis, the way you can best model the real-world situation with respect to the surface radii is to first assign a power fit to the testplate in units of fringes. You will then also assign a radius tolerance, which in this case means the *accuracy* to which the testplate radius is known to have been measured. This is a measurement of accuracy or capability tolerance, and it should be discussed with your lens manufacturer. And if you really want to assure that your system works well, have the shop remeasure the testplates you have matched and incorporate these newly measured results into your design. After all, equipment and techniques constantly improve, so why depend on a testplate radius measurement from 20 years ago?

On a different, yet related matter, for extremely high-precision systems the designer often incorporates the measured refractive indices of the glasses into the design. This is called a *melt design*, and is yet a further assurance that the design will perform as predicted. As with testplates, glass refractive indices can be measured more accurately than they can be manufactured.

How to Tolerance an Optical System

The basic procedure for tolerancing an optical system is shown in outline form in Fig. 16.1 and is described as follows:

1. We first assign viable tolerances to all toleranceable parameters within our system. This includes all optical as well as mechanical components. This first candidate set of tolerances should be reasonably achievable at a rational cost. If we know we have a very sensitive system and/or we require a very high level of performance, a somewhat tighter set of tolerances should be used; conversely if we have an insensitive system or a poorer level of performance, then looser tolerances can be used.

For an optical system with a reasonable level of performance, a candidate tolerance set may be derived from Table 16.1.

Note in Table 16.1 that the radius tolerance is listed as testplate measurement accuracy. This is the accuracy to which the testplates have

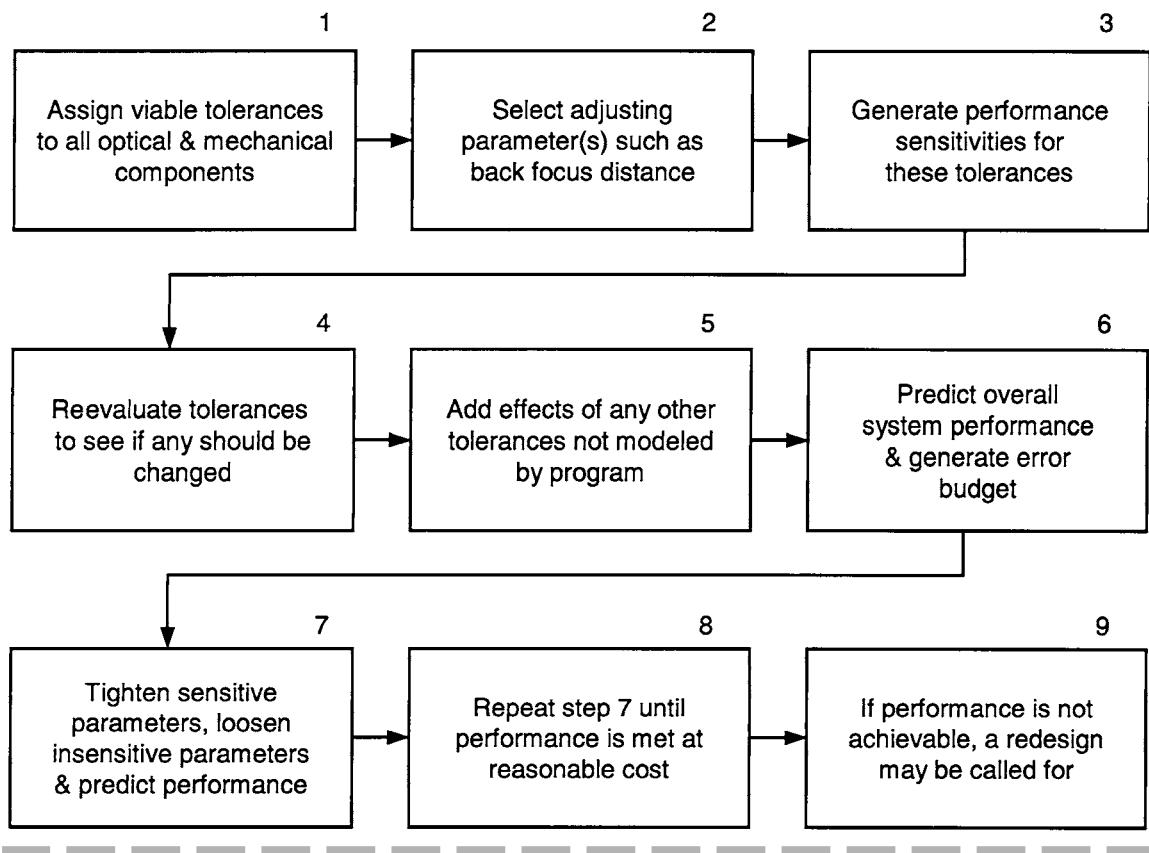


Figure 16.1
Tolerancing Procedure

TABLE 16.1

Candidate Tolerances for a Reasonable Performance Lens

Parameter	Tolerance	Parameter	Tolerance
Radius	Testplate measurement accuracy	Tilt	0.05-mm TIR
Power fit to testplate	Three fringes	Decenter	± 0.05 mm
Surface irregularity	One fringe	Refractive index	± 0.001
Thickness	± 0.05 mm	Abbe number	$\pm 0.8\%$
Air space	± 0.05 mm	Glass inhomogeneity	± 0.0001
Wedge/centration	0.025-mm TIR		

been measured at the shop. Using a value of ± 0.01 mm is probably a reasonable assumption, but check with your shop to be sure. For long radii, a better assumption might be the radius change corresponding to ± 0.25 wave of sag at the outer periphery of the element clear aperture, or the testplate, whichever is smaller. This is because for long radii testplates the effective *f*/number of the light cone from the center of curvature of the surface is high, thus yielding a sizable depth of focus, which ultimately impacts the ability of determining the radius.

2. We then generate performance degradation sensitivities for all toleranceable parameters within the system using the tolerance routine in our lens design computer program. In other words, each and every fabrication, assembly, and/or alignment-related tolerance on all components is evaluated and sensitivities are determined. The tolerance forms include radius, power fit to testplate, surface irregularity, element thickness, airspace, element wedge, element tilt, element decentration, refractive index, Abbe number, and glass inhomogeneity. Other specific factors, such as the effects to performance due to the thermal environment, may also need to be included here.
3. As part of the preceding tolerance sensitivity analysis, we need to select the appropriate *adjusting parameter* or *parameters*. An adjusting parameter is an adjustment which you plan to allow for during the final lens assembly and testing. Back focus is the most common adjusting parameter and will almost always be used as a final adjustment during final system assembly. There can, in principle, be other adjusting parameters ranging from airspaces to tilts and decentrations. More on adjusting parameters later.
4. We now generate the performance sensitivities for this initial set of candidate tolerances. Thus, if our performance criteria were the MTF at 30 line pair/mm, we would determine the drop in MTF at this spatial frequency for each toleranceable parameter within the system.
5. We now go back and look carefully at each tolerance to determine if it should be changed from a manufacturing or assembly point of view. For example, larger airspaces may require somewhat looser tolerances. This is the all-important time to talk with your optical shop and your mechanical designer as well as your machine shop to reach a mutual understanding of the optical components and the mechanical design with respect to the anticipated levels of tolerances, which will affect directly the lens elements.

6. We now add the performance degradations from any other effects, which may not have been covered by our computer model. These may include atmospheric turbulence, surface irregularities that may not have been modeled in the computer analysis, and/or other effects. The net result is to predict the expected level of optical performance based on the assumed tolerance set.
7. We then predict the overall system performance and generate a performance error budget.
8. We now proceed to tighten any sensitive parameters and loosen insensitive ones and, again, predict performance. Here you need to look carefully at the effect on your performance metric such as MTF for each toleranced parameter to determine its effect. You will sometimes find that only a small subset of the overall tolerance set is sensitive, and the majority of the tolerances are insensitive. In these situations, tightening only a small number of tolerances can have a big payoff in improved performance. It is important to review the tolerances you intend to tighten with your optical shop and/or machine shop.
9. Step 8 is repeated as necessary until we meet the performance goal at a reasonable level of cost. The results of step 8 comprise the final set of manufacturing and assembly tolerances which, if applied and adhered to, will result in a system which meets your performance goals and objectives at a reasonable cost.
10. If the required performance is not achievable at a reasonable level of cost, you may need to return to your initial specifications to see which ones you may be able to relax. Also, a redesign may be called for with the specific goal of loosening the manufacturing tolerances.

How Image Degradations from Different Tolerances Are Summed

When you compute tolerance sensitivities as outlined earlier, you will ultimately need to predict the net system performance. Ideally, this should represent the predicted level of performance for some reasonably high percentage of manufactured systems such as a 95% cumulative probability or confidence level. But how do we add performance degradations in order to accurately predict performance? In other words, if we have, for example, 10 tolerances and each one degrades the performance

by a different amount, how do we use these data to predict the net result of assembling a large number of systems?

One common method of adding degradations is known as *RSS addition*, which stands for root sum square. In this method, we take the *square root of the sum of the squares* of each of the individual degradations, and if the degradations are of the same form, this should lead to a 95% confidence level. The degradations could be a peak-to-valley optical path difference (OPD), RMS OPD drop in MTF or some other criteria. While this can all be done, the criteria are not all mathematically correct, as will be discussed later.

To illustrate RSS addition and how powerful the method is, let us assume that we have a stack of pennies, and each penny has thickness $T \pm 0.1$ mm. If we have 100 pennies, the predicted thickness of the stack of 100 pennies is $100 \cdot T \pm 10$ mm worst case or $\sqrt{100} T \pm 1$ mm for a 95% confidence based on an RSS addition, since $\sqrt{0.1^2 \times 100} = 1$. The tolerance on the thickness of a penny is of a simple form in that the penny is either thicker or thinner than nominal; in other words, the degradations are all of the same type—they are all thickness errors.

In lenses, the degradations to performance are of different forms and are caused by different types of parameter errors. For example, performance degradation introduced by center thickness and airspace errors will likely introduce defocus plus spherical aberration on axis. On the other hand, element wedges, tilts, and decentrations will likely introduce coma and/or astigmatism on axis and across the field of view. The question now becomes: can we use an RSS addition to predict the net expected performance due to axial spacing errors as well as asymmetrical errors such as decentrations? The answer is a qualified “yes,” we can indeed RSS performance degradations caused by unlike errors (such as thickness errors and decentrations), but the more important question is whether the results are accurate and meaningful. The bottom line is that RSS addition simply does not handle properly the mix of aberrations that we often encounter in a lens system. For certain cases, RSS addition is quite appropriate and valid, and one such case is in computing the expected refocusing required following assembly. If we take the RSS of the refocusing required for each and every individual tolerance, the result should be valid since all of the refocusing is of the same form.

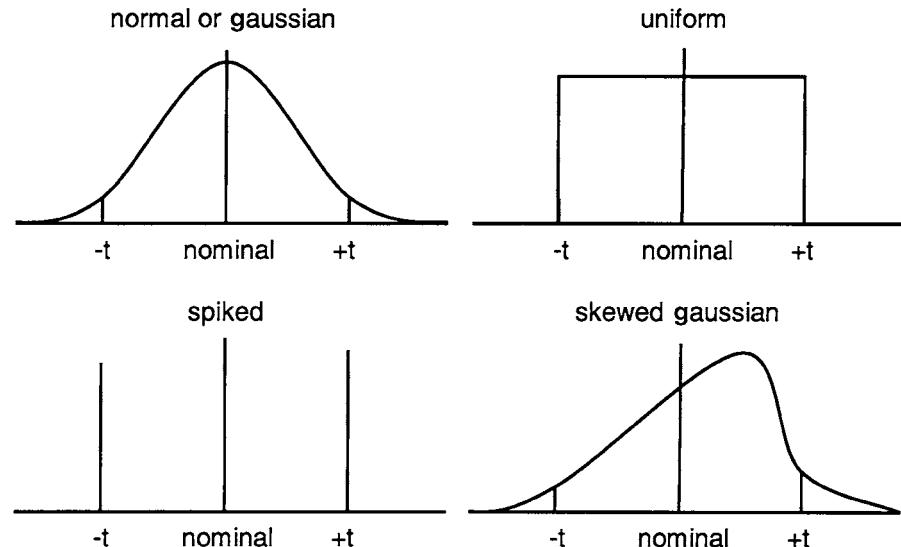
Fortunately, there is another approach known as *Monte Carlo tolerancing*, which is a method of simulating the performance statistics of a lens system in a high level of production. In a Monte Carlo simulation, tolerances are assigned to all toleranceable parameters along with a likely

probability function such as normal, uniform, spiked, or a skewed gaussian, as shown in Fig. 16.2. While a normal distribution seems to be the best model, we often find that tolerances sometimes tend to end up toward one end of the allowable range. Element center thickness is a good example, as the optician often leaves extra thickness just in case a scratch occurs and the element needs to be reground. Thus, the probability distribution for CTs is a skewed gaussian distribution with the highest probability on the thick side of nominal as shown in Fig. 16.2.

The following is the procedure used in computing a Monte Carlo tolerance analysis:

- Every parameter is independently and randomly perturbed according to its assigned tolerance, based on a likely probability distribution. At this point, we have created a random lens assembly on the computer, based on our current set of tolerances.
- The performance is now computed after applying any compensators such as refocusing. The system model at this point is a simulation of a single manufactured system with its tolerances randomly distributed as discussed previously.
- The preceding process is repeated 25 or more times, and for each Monte Carlo sample, we have, in effect, a simulated system that has been manufactured. We can use the resulting output to compute

Figure 16.2
Tolerance Distribution Models



the level of performance versus the cumulative probability. With these results, we can easily determine the level of performance for, say, 90 or 95% cumulative probability of occurrence.

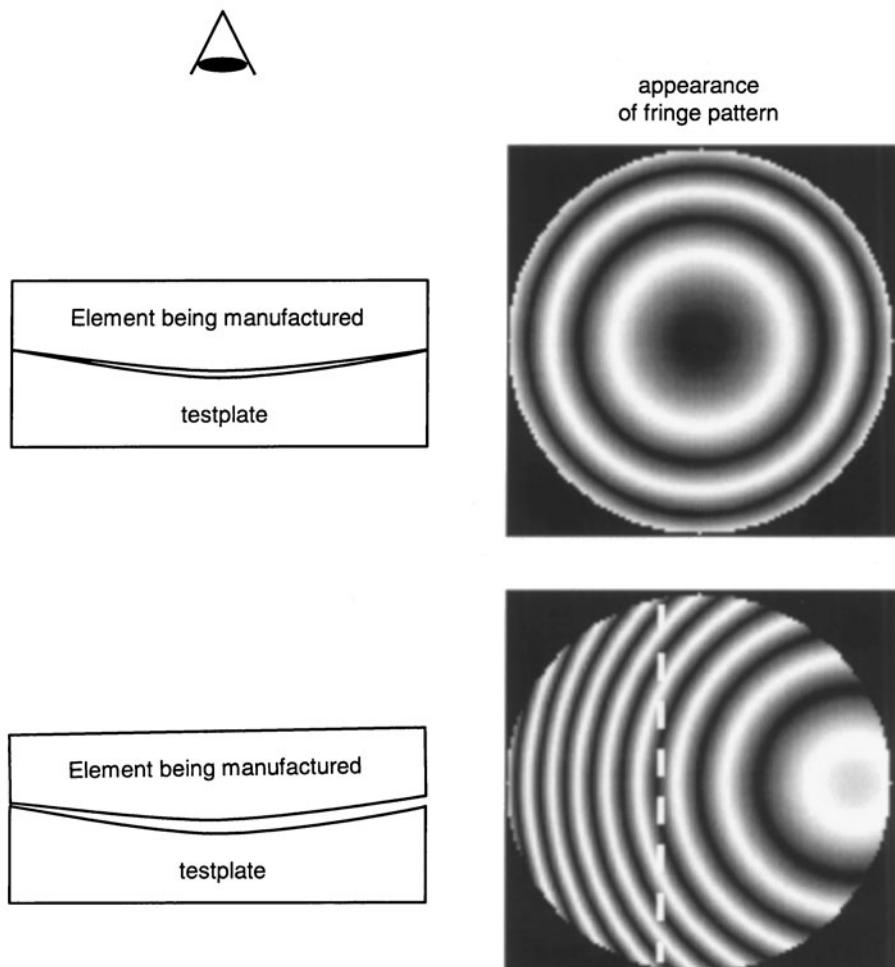
The beauty of the Monte Carlo approach is that it is fully valid regardless of the relative nature, mix, or form of the aberrations because we are simulating a manufacturing environment. Each and every Monte Carlo sample is, in effect, a new and different manufactured system.

Forms of Tolerances

We have discussed the various tolerances, along with ways and means for adding the performance degradations. Next we will show the various forms of the tolerance parameters:

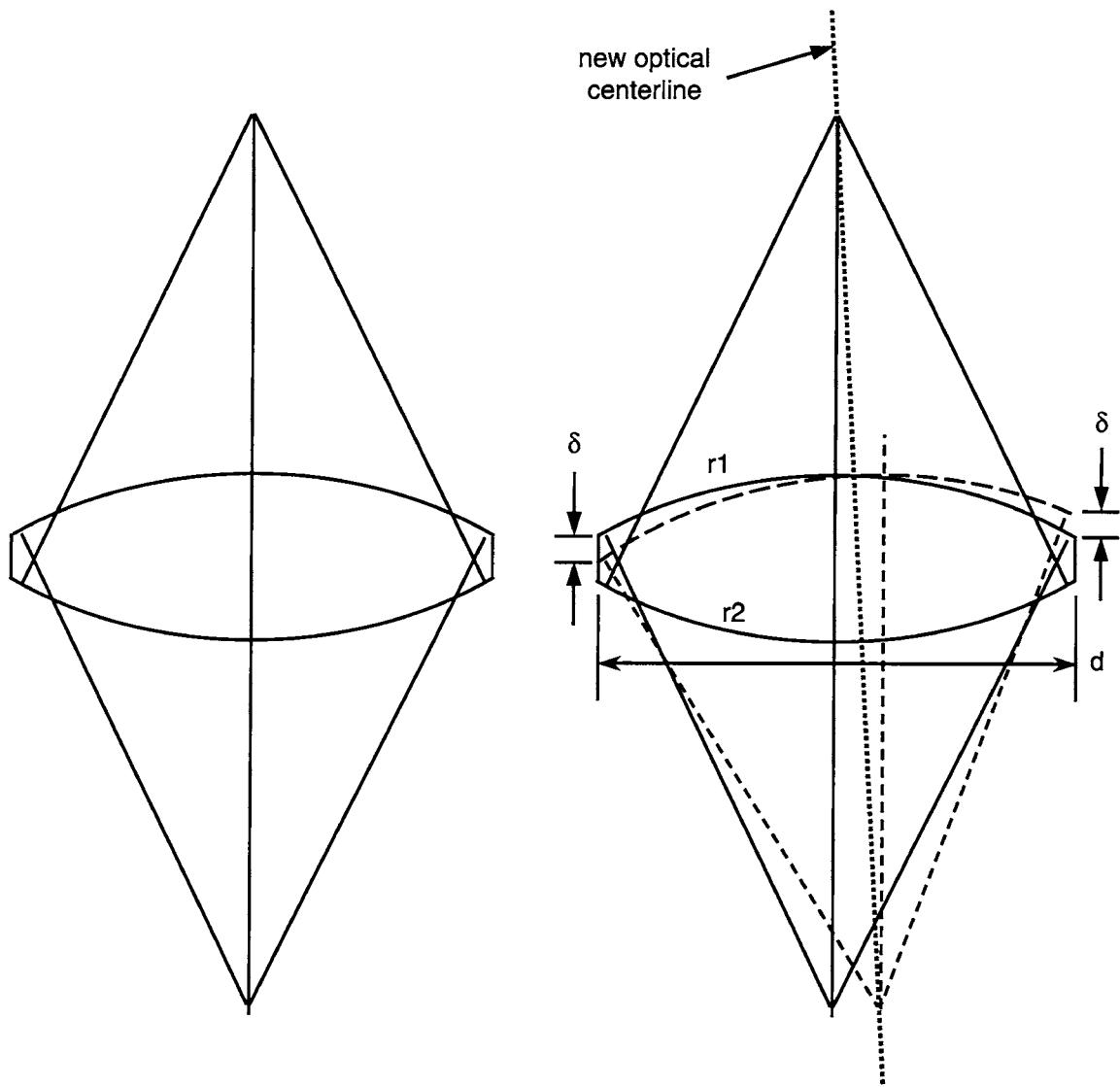
- *Symmetrical errors* relating to fabrication and assembly include radius, power fit to testplate, thickness of elements, airspaces, refractive index, and Abbe number. Figure 16.3 shows how a testplate is used to characterize a lens element in manufacturing. The testplate is used as a standard of a known radius to high precision. The surface under test is placed, or “nested,” into the testplate as shown, and every time the airspace or gap between it and the surface under test changes by $\lambda/2$ we see one full fringe. In the example shown, we see approximately two rings, or fringes, which means that approximately 1.0 wave of optical power or mismatch exists between the testplate and the surface under test. If the gap between the testplate and the surface under test is rotationally symmetrical, the interference pattern is indeed round, with circularly symmetric rings, as shown in the upper part of Fig. 16.3. If, on the other hand, we tilt the surface under test (this is the same as moving the ring center way off the part as shown), we see curved fringes as in the lower part of Fig. 16.3. The dashed vertical line is a reference to tell us how many fringes of power we have, and, as you can see, we have about two fringes of power.
- *Asymmetrical errors* in assembly and alignment include element wedge, element decentration, element tilt, surface irregularity, and inhomogeneity of refractive index. In effect, an element with a wedge is really the same as an element with its optical centerline tilted with respect to its mechanical centerline. Further, the element

Figure 16.3
Power Fit to Testplate



has an *edge thickness difference* as we rotate the element around.

Consider Fig. 16.4 where on the left we show a nominal perfect lens element. The two centers of curvature, when connected, represent a line straight through the mechanical centerline of the element. This line is, by definition, the optical axis of the element. On the right is an element with a severe tilt on its top surface. Note that this surface tilt or element wedge results in an edge thickness difference from left to right (the edge is thinner on the left, thicker on the right). The wedge, in radians, is the edge thickness difference, 2δ , divided by the element diameter, d . Thus, the wedge, in radians, equals $2\delta/d$.

**Figure 16.4**

Left: Nominal Element; Right: Wedged Element

As we have shown, a lens element with spherical surfaces will, by definition, have a single optical axis, and it is the process of lens centering which will bring the mechanical edge of the element to be concentric with the optical axis of the lens. Figure 16.5 shows two mechanical methods. In Fig. 16.5a the lens is located on a precision spindle so that its lower surface runs true. Spinning the element will result in a large total indicator runout (TIR) reading on the dial indicator. Moving the element to the left in its shown rotational position will enable the top surface to run true, and the element is then edged using a diamond wheel. Figure 16.5b shows a method of centering called "cupping," in which the upper and lower rings (similar to the edge on a cup) will only fully contact the lens surfaces when the two surfaces are on a common and vertical axis. When this alignment condition is achieved, the element is edged to the proper diameter.

In many of today's optical shops techniques using HeNe lasers are often used to aid in centering lens elements. With one surface running

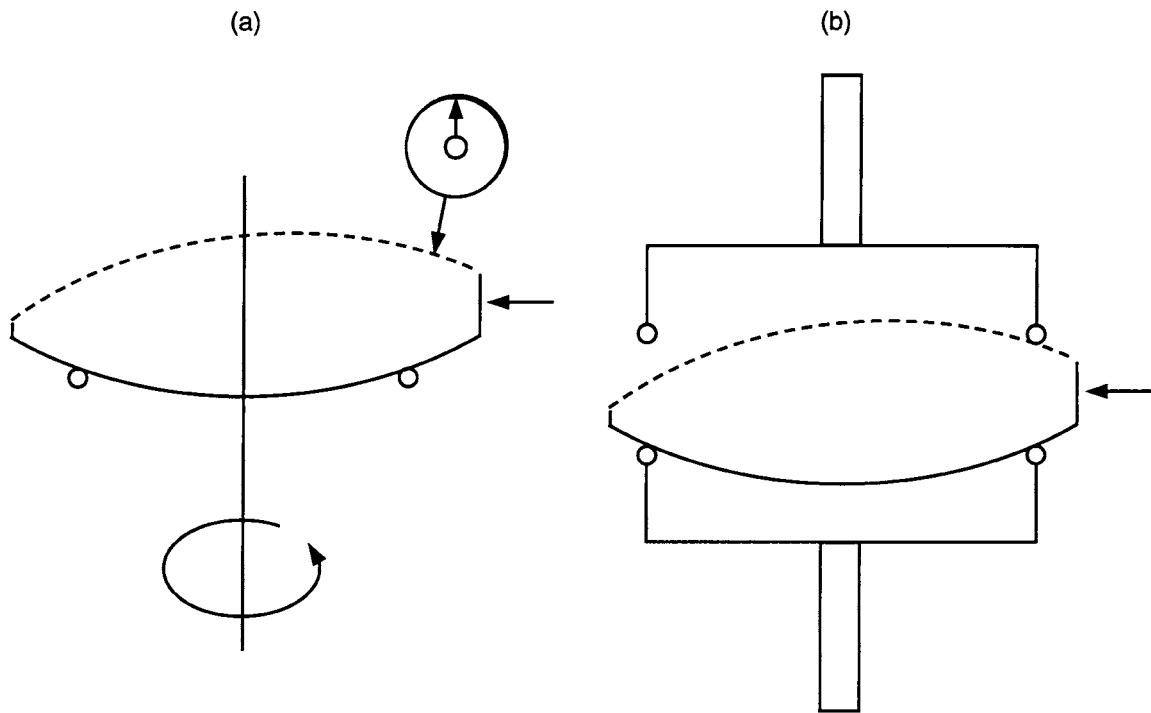


Figure 16.5
Left: Centering on a Spindle; Right: by Cupping

true, the laser beam is either reflected off of the other surface or transmitted through the element, while the element is being rotated. The laser beam will nutate, or "wobble," until the element is properly centered. By using sensors such as CCDs or quad cells, extremely accurate indications of the element centration can be made.

Element decentration can be either a simple lateral decenter (up and down) or it can be a "roll" whereby the element maintains contact with a housing seat, as shown in Fig. 16.6. Note that while the net effect is quite similar, the two decentration models are actually quite different. In the roll situation the left-hand radius, which is in perfect contact with the housing seat, ends up perfectly aligned, with the surface tilt occurring only on the right-hand surface. Many of the computer software packages can model both pure decentrations as well as roll.

Finally, element tilt is self-explanatory, and it is expressed either as a total indicator runout (TIR) or sometimes as the tilt, in minutes of arc.

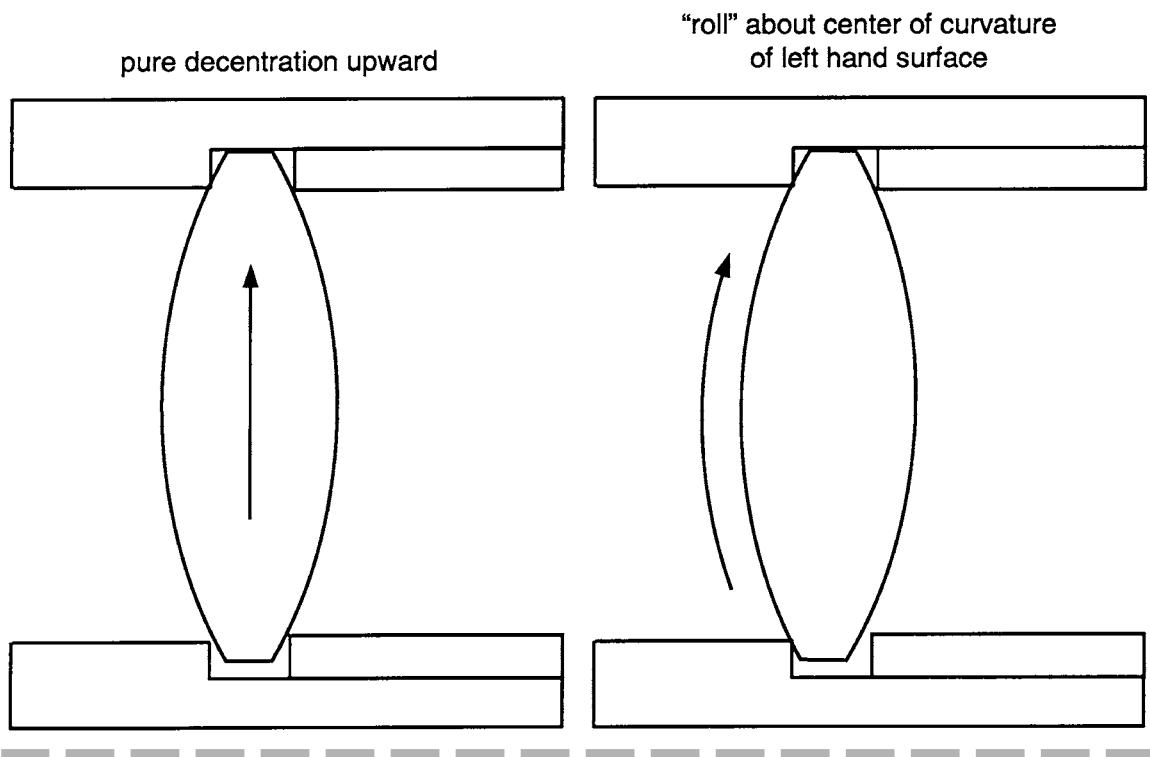


Figure 16.6
Element Decentration Models

Since 1 min of arc is 0.0003 rad, we simply multiply the tilt, in radians, by the diameter to derive the TIR.

Adjusting Parameters

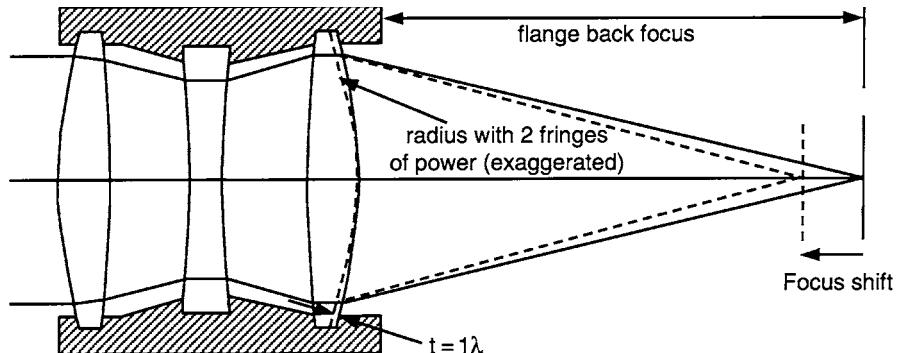
Adjusting parameters, sometimes called *compensators*, are those parameters used to optimize the optical performance in the laboratory or on the assembly line at some time during the final lens assembly and testing.

The most common adjusting parameter is refocusing during the final assembly and testing procedure. Let's consider the manufacturing of a 35-mm camera lens or a similar high-quality lens for a CCD camera. The hypothetical example shown in Fig. 16.7 illustrates the situation. The lens will be mounted to the camera using a bayonet or similar accurate mounting methodology, and the distance from the rear flange to the image plane (the film or the CCD chip) is a tightly held mechanical dimension. Needless to say it is imperative that when the completed lens is fastened to the flange on the camera body, the image of an infinite object is in perfect focus on the film or CCD when the lens is focused at infinity.

Referring to Fig. 16.7, we show a nominal lens imaging onto a sensor at the nominal flange back-focal distance. We also show in dashed lines the last radius of the lens with a steeper or more powerful radius than nominal by two fringes of power. The effect of this is to move the image inward, and the image will be out of focus on the sensor.

The optical path difference (OPD) = $(n - 1)t$, where n = the refractive index and t = the separation between the nominal surface and the manufactured surface at the edge of the aperture as shown. If we have two

Figure 16.7
Refocusing As an
Adjusting Parameter



fringes of power; this is due to one wave of surface error, or sag, at the edge of the element. And since $OPD = (n - 1)t$, and $t = 1.0\lambda$, the $OPD = 0.5\lambda$ peak to valley. If the flange back-focus distance is absolutely correct for the nominal camera and for the nominal lens assembly, the image will be out of focus by 0.5λ peak to valley, which is a factor of 2 from the Rayleigh criteria, due *only* to the last radius being two fringes of power from nominal!

The entire lens in this simplified example of course has six radii as well as the two airspaces, element thicknesses, and other tolerances, all of which will further contribute to defocus errors. If we RSS the effect of the six radii being two fringes of power each, we find that the final image could be out of focus by 1.22 waves of defocus, which is of major significance. What this means is that if we consider only the power fit to testplate, then with a tolerance of two fringes per surface, we predict a focus error of 1.22 waves of defocus with a 95% confidence level. This is about 5 times the Rayleigh criteria, and the imagery will be poor relative to the diffraction limit. We *must* refocus the lens, and this is accomplished by using an adjusting parameter, in this case the back-focus distance, during final assembly.

It is not uncommon to sometimes use an element decentration as an adjusting parameter. This is done when the system is quite sensitive to what is sometimes called *axial coma*, or coma which occurs on axis (and carries somewhat uniformly across the field of view) due to asymmetrical tolerances such as element wedge, tilt, and decentration. By allowing a strategically located element within the system to be adjusted in *X* and *Y* during the final assembly and testing, you can often cancel the coma introduced by all of the other tolerances within the system. Not only will this permit you to produce a system which otherwise may not work properly, but you may be able to relax some of your other tolerances from their otherwise tight levels, thus lowering the cost and enhancing the producibility. The trade-off often results in producing a system with extremely tight tolerances versus looser tolerances and one adjustment made during final assembly and testing.

In large field-of-view projection optics, it is sometimes imperative to have a lens or a group of lenses that are adjusted in centration during final assembly of the system in order to relax the tolerances on the other components, and have a reasonable cost to the system.

In microscope objective manufacturing a common test is the “star test” discussed in Chap. 15. In some high-precision shops, if the objective shows axial coma, it is sometimes tapped or “banged” lightly on a table and retested one or more times until the performance is met. While somewhat of a “brute force” method of compensator usage, it does work.

Typical Tolerances for Various Cost Models

There have been several papers presented over the years showing the effect on cost associated with various levels of tolerances. One of these, presented by John Plummer, is shown in Fig. 16.8. Note that there is a similar updated table in Chap. 17. You may find it interesting to compare

diameter tolerance in millimeters	± 0.1 100	± 0.05 100	± 0.025 103	± 0.0125 115	± 0.0075 150		
center thickness tolerance in mm	± 0.2 100	± 0.1 105	± 0.05 115	± 0.025 150	± 0.0125 300		
stain charac. of the glass	0 100	1 100	2 103	3 110	4 150	5 250	5+ 500
# of lenses per block 100	25 105	18 115	11 130	6 175	3 300	1	
eccen. Toler. in light dev.	6 min. 100	3 min. 103	2 min. 108	1 min. 115	30 sec. 140	15 sec. 200	
Figure tolerance in λ (pow/irreg.)	10-5 100	5-2 105	3-1 120	2 1/2 140	2 1/4 175	1 1/8 300	
Dia. to thick. ratio (fig. 3.1)	9-1 100	15-1 120	20-1 150	30-1 200	40-1 300	50-1 500	
Beauty defects (MIL-C-13830A)	80-50 100	60-4 110	40-30 125	20-10 175	10-5 350		
Raw glass cost in 1000 lb lots	\$3.00 100	\$5.00 108	\$8.00 115	\$15.00 125	\$25.00 135	\$50.00 200	\$100.00 350
Coating speci- fications	unctd. 100	Mg. Fl. 115	3-4 layer 150	>4 layers 200-500			

Figure 16.8

Relative Cost of Manufacturing As a Function of the Level of Tolerance

the two sets of data. Plummer's company, at the time, was primarily involved in high production of reasonable quality optics which we might call "riflescope quality" in other words not ultratight tolerances and not "loosey goosey" either. For 10 manufacturing parameters, the level of tolerance is shown on the first line and the relative cost on the second line for various levels of tolerances. Let's look at several examples:

- The standard diameter tolerance is ± 0.05 mm. If we require ± 0.025 , the cost increase is only about 3%; however, if we need ± 0.01 , then the cost increase is about 25 to 30%.
- Meeting tight tolerances for element thicknesses is difficult and costly. Achieving a tolerance of ± 0.05 costs 15% above the standard level of ± 0.2 , and achieving ± 0.025 will cost an additional 50%.
- Even stain characteristics on glass will affect cost. The reason for this is that a glass type with a 5 stain will acquire a stain within minutes, and for this reason the elements must be sent almost immediately into the vacuum chamber for coating after they have been polished. This, of course, affects the workflow in the shop and for this reason it becomes costly.

You should take the time to read through the table in Fig. 16.8 and become familiar with the various tolerances and their relative cost versus quality trade-offs. More in-depth material relating to optical manufacturing and tolerances is presented in Chap. 18.

Example of Tolerance Analysis

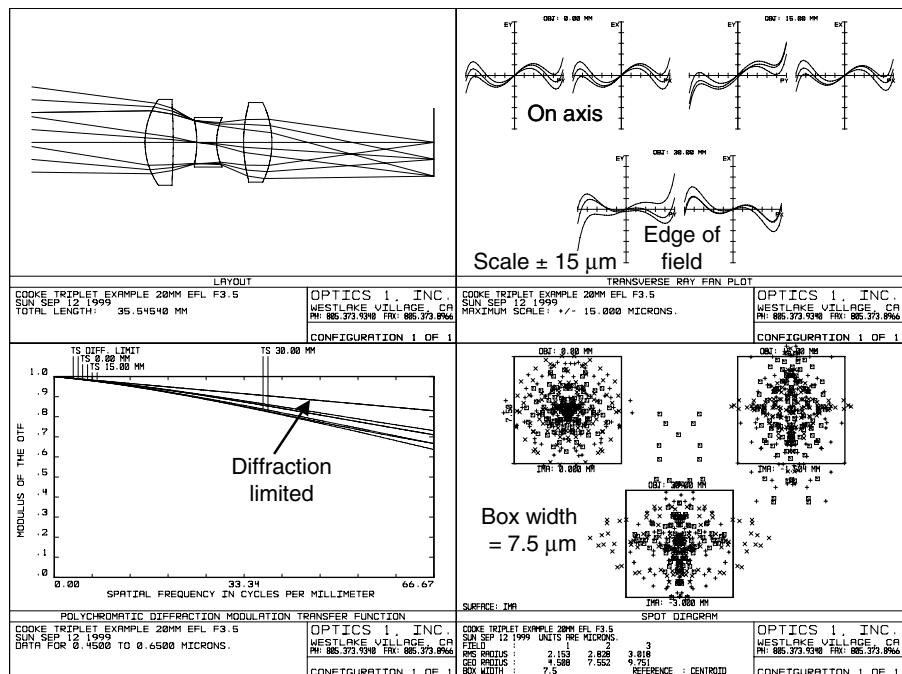
In order to best show how to tolerance a lens system, we will go through the tolerances for a $10\times$ reduction lens Cooke triplet for a machine vision application. Our basic specifications are as follows:

Parameter	Specification
Object distance	200 mm
Object full diagonal	60 mm
Magnification	$0.1\times$
Image full diagonal	6 mm
f/number	f/3.5 at used conjugate

Parameter	Specification
Focal length	≈ 20 mm
Full field of view	$\approx 17.06^\circ$
Spectral band	550 to 650 nm, uniform weights
Sensor	1/3-in CCD
Number of pixels	640×480
Pixel pitch at image	$7.5 \times 7.5 \mu\text{m}$
Pixel pitch at object	$75 \mu\text{m}$
Nyquist frequency at image	66.67 line pair/mm
Nyquist frequency at object	6.667 line pair/mm
MTF spec at image	>0.3 at Nyquist

Figure 16.9 shows the layout and performance of the final design for this lens. Note that the design is quite close to diffraction limited as evidenced from the MTF. The geometrically based spot diagrams

Figure 16.9
Cooke Triplet for
Machine Vision
Application



show square boxes measuring $7.5 \times 7.5 \mu\text{m}$, which is one pixel at the CCD sensor. Figure 16.10 is a listing of the specifications and prescription data for the lens in case you want to set it up and work with it.

We will now apply a standard set of manufacturing tolerances to the lens. We show in the following the resulting tolerance sensitivities. In order to conserve space, data will be shown for the central element only (this is for most of the tolerances, the most sensitive). In the data, we see in the "Field" column "All," which is the average of all of the fields of view. Below that are each of the individual fields (field 1 is on axis, 2 and 3 are $\pm 70\%$ of the full diagonal, and 4 and 5 are \pm full field). The "MTF" columns are the MTF after applying the tolerance and refocusing for compensation (by the "Change in Focus" value), and the "Change" columns are the change in MTF

Figure 16.10
Prescription of Cooke
Triplet for Machine
Vision Application

System/Prescription Data					
GENERAL LENS DATA:					
Surfaces :	8				
Stop :	4				
System Aperture : Image Space F/#	= 3.5				
Eff. Focal Len. :	19.00051	(in image space)			
Back Focal Len. :	12.4271				
Total Track :	35.48153				
Image Space F/# :	3.5				
Stop Radius :	1.777314				
Parax. Ima. Hgt.:	3.018941				
Parax. Mag. :	-0.1006314				
Entr. Pup. Dia. :	5.428717				
Entr. Pup. Pos. :	15.4601				
Exit Pupil Dia. :	6.196183				
Exit Pupil Pos. :	-23.54149				
Field Type :	Object height in Millimeters				
Maximum Field :	30				
Primary Wave :	0.55				
Lens Units :	Millimeters				
SURFACE DATA SUMMARY:					
Surf	Type	Radius	Thickness	Glass	Diameter
OBJ	STANDARD	Infinity	190		60
1	STANDARD	Infinity	10		0
2	STANDARD	6.572511	2.5	SK5	7.303856
3	STANDARD	-51.05087	2.041158		6.332172
STO	STANDARD	-8.89512	1.75	SF2	3.696617
5	STANDARD	5.048865	2.408424		4.172604
6	STANDARD	12.99962	2.5	SK5	6.31239
7	STANDARD	-8.044117	14.28195		6.824242
IMA	STANDARD	Infinity			6.020948

In the following data, the tolerance designations are:

- TFRN is the number of fringes of power fit to the testplate.
- TTHI is the airspace or element thickness, in millimeters.
- TEDY is the decentration of an element, in millimeters.
- TETY is the tolerance of element tilt, in degrees.
- TIRY is the tolerance on the total indicator runout or surface tilt for wedge or element centration.
- TIRR is the tolerance on surface irregularity in fringes.

Type	Sf1	Sf2	Field	Minimum			Maximum		
				Value	MTF	Change	Value	MTF	Change
TFRN	3	All	-4.00	0.713	-0.027	4.00	0.710	-0.031	4 fringes power to testplate
	1		0.739	-0.023		0.724	-0.038		
	2		0.659	-0.060		0.726	0.006		
	3		0.659	-0.060		0.726	0.006		
	4		0.751	0.008		0.681	-0.061		
	5		0.751	0.008		0.681	-0.061		
Change in Focus :				0.126			-0.124		
TFRN	4	All	-4.00	0.732	-0.008	4.00	0.727	-0.013	4 fringes power to testplate
	1		0.754	-0.008		0.759	-0.003		
	2		0.731	0.011		0.688	-0.031		
	3		0.731	0.011		0.688	-0.031		
	4		0.714	-0.028		0.740	-0.002		
	5		0.714	-0.028		0.740	-0.002		
Change in Focus :				-0.096			0.097		
TTHI	3	4	All	-0.050	0.720	-0.020	0.05	0.720	-0.021 thickness
	1		0.764	0.001		0.761	-0.001		
	2		0.746	0.026		0.679	-0.040		
	3		0.746	0.026		0.679	-0.040		
	4		0.662	-0.080		0.726	-0.016		
	5		0.662	-0.080		0.726	-0.016		
Change in Focus :				0.045			-0.045		
TTHI	4	6	All	-0.050	0.729	-0.011	0.050	0.732	-0.008 airspace
	1		0.763	0.000		0.762	-0.000		
	2		0.694	-0.025		0.736	0.015		
	3		0.694	-0.025		0.736	0.015		
	4		0.736	-0.006		0.702	-0.040		
	5		0.736	-0.006		0.702	-0.040		
Change in Focus :				-0.042			0.042		
TEDX	3	4	All	-0.025	0.607	-0.134	0.025	0.607	-0.134 element
	1		0.633	-0.129		0.633	-0.129		
	2		0.587	-0.132		0.587	-0.132		
	3		0.587	-0.132		0.587	-0.132		
	4		0.602	-0.140		0.602	-0.140		
	5		0.602	-0.140		0.602	-0.140		
Change in Focus :				0.000			0.000		
TEDY	3	4	All	-0.025	0.589	-0.151	0.025	0.589	-0.151 element
	1		0.633	-0.129		0.633	-0.129		
	2		0.550	-0.169		0.607	-0.112		
	3		0.607	-0.112		0.550	-0.169		

				4	0.563	-0.179		0.559	-0.183	
				5	0.559	-0.183		0.563	-0.179	
					:	0.000			0.000	
				Change in Focus						
TETX	3	4	All	-0.150	0.584	-0.157	0.150	0.584	-0.157	element tilt x
				1	0.757	-0.005		0.757	-0.005	
				2	0.703	-0.016		0.477	-0.242	
				3	0.477	-0.242		0.703	-0.016	
				4	0.417	-0.325		0.532	-0.210	
				5	0.532	-0.210		0.417	-0.325	
					:	0.000			0.000	
TETY	3	4	All	-0.15	0.661	-0.080	0.150	0.661	-0.080	element tilt y
				1	0.757	-0.005		0.757	-0.005	
				2	0.648	-0.071		0.648	-0.071	
				3	0.648	-0.071		0.648	-0.071	
				4	0.598	-0.144		0.598	-0.144	
				5	0.598	-0.144		0.598	-0.144	
					:	0.000			0.000	
TIRX	3	All	-0.015	0.566	-0.174	0.015	0.566	-0.174	wedge TIR x	
				1	0.638	-0.124		0.638	-0.124	
				2	0.552	-0.167		0.552	-0.167	
				3	0.552	-0.167		0.552	-0.167	
				4	0.518	-0.224		0.518	-0.224	
				5	0.518	-0.224		0.518	-0.224	
					:	0.000			0.000	
TIRY	3	All	-0.015	0.570	-0.170	0.015	0.570	-0.170	wedge TIR y	
				1	0.638	-0.124		0.638	-0.124	
				2	0.629	-0.090		0.508	-0.211	
				3	0.508	-0.211		0.629	-0.090	
				4	0.482	-0.260		0.555	-0.186	
				5	0.555	-0.186		0.482	-0.260	
					:	0.000			0.000	
TIRX	4	All	-0.015	0.683	-0.058	0.015	0.683	-0.058	wedge TIR x	
				1	0.736	-0.026		0.736	-0.026	
				2	0.664	-0.055		0.664	-0.055	
				3	0.664	-0.055		0.664	-0.055	
				4	0.654	-0.088		0.654	-0.088	
				5	0.654	-0.088		0.654	-0.088	
					:	0.000			0.000	
TIRY	4	All	-0.015	0.627	-0.114	0.015	0.627	-0.114	wedge TIR y	
				1	0.736	-0.02		0.736	-0.026	
				2	0.708	-0.011		0.539	-0.180	
				3	0.539	-0.180		0.708	-0.011	
				4	0.507	-0.235		0.604	-0.138	
				5	0.604	-0.138		0.507	-0.235	
					:	0.000			0.000	
TIRR	3	All	-1.000	0.636	-0.104	1.000	0.652	-0.089	surface	
				1	0.669	-0.093		0.651	-0.111	irregularity
				2	0.554	-0.165		0.721	0.001	
				3	0.554	-0.165		0.721	0.001	
				4	0.704	-0.038		0.595	-0.147	
				5	0.704	-0.038		0.595	-0.147	
					:	0.000			0.000	
TIRR	4	All	-1.00	0.709	-0.032	1.000	0.691	-0.050	surface	
				1	0.719	-0.043		0.730	-0.032	irregularity
				2	0.748	0.028		0.624	-0.095	
				3	0.748	0.028		0.624	-0.095	
				4	0.665	-0.077		0.730	-0.012	
				5	0.665	-0.077		0.730	-0.012	
					:	0.000			0.000	

The preceding are the tolerance sensitivities for the central element. There are many numbers... what do they all mean?

At this point, you should look over all of the sensitivities to see if any are especially sensitive. For example, decentration of the central element by 0.025 mm drops the average MTF about 0.134 for an x decentration (in and out of the figure) and 0.151 for a y decentration. The effect over the field is reasonably uniform (this is not always the case). A less sensitive tolerance is the four fringes of power on surface 4 (the rear of the second element), where a maximum MTF drop of only 0.013 average (0.031 maximum drop) has resulted. We should be able to loosen this tolerance if our shop feels it is worthwhile.

In Fig. 16.11 we show in a graphical form those parameters which drop the average MTF over the field of view by 0.02 or more for our lens. There are many tolerances not even represented in the data. Note that there are only about five to six tolerances that are most sensitive, and these are primarily tilts, decentrations, and wedges. The most sensitive tolerances are for element 2, as discussed earlier.

The best measure of the overall lens performance and the most reliable means for predicting performance is via the Monte Carlo analysis. Here we computed 20 Monte Carlo samples and the resulting statistics are shown in Table 16.2. Recall that each Monte Carlo sample is, in effect, a simulated fabricated system. Each individual parameter is changed according to a normal probability distribution between its minimum and maximum values.

Figure 16.11
All Cooke Triplet Tolerances which Drop the MTF by >0.02

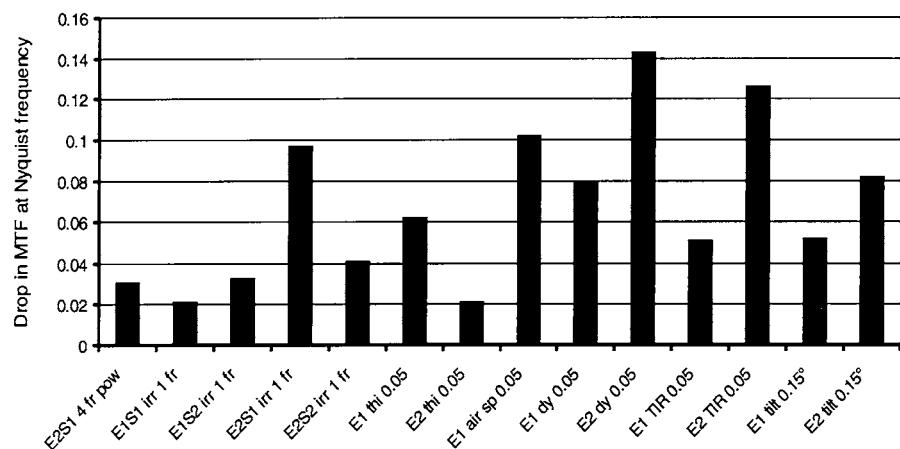


TABLE 16.2

Monte Carlo Results

	Average	Field				
		1	2	3	4	5
Nominal	0.741	0.762	0.720	0.720	0.742	0.742
Best	0.602	0.760	0.664	0.649	0.616	0.641
Worst	0.326	0.360	0.297	0.327	0.196	0.165
Mean	0.491	0.585	0.484	0.498	0.444	0.410
Standard deviation	0.068	0.108	0.101	0.086	0.105	0.107

Compensator Statistics	
Change in back focus:	
Minimum	-0.232282
Maximum	0.222737
Mean	0.031034
Standard deviation	0.138210

Ninety percent of Monte Carlo lenses have an MTF above 0.306.

Fifty percent of Monte Carlo lenses have an MTF above 0.393.

Ten percent of Monte Carlo lenses have an MTF above 0.556.

The Monte Carlo results show that 90% of our lenses should have an MTF at 66.7 line pairs/mm of 0.306 or better. This just meets our MTF goal of 0.3 or better! It still may be beneficial to review each tolerance to see if it can be loosened, and the most sensitive tolerances should be tightened if possible. Once these changes have been made, another Monte Carlo analysis is in order as this is our best way of modeling the predicted lens performance.

It appears that 20 Monte Carlo samples is a small number to represent accurately the statistics of assembling lenses in production. So, we ran 500 Monte Carlo samples with the same tolerances. Here 90% of the lenses are predicted to have an MTF at Nyquist greater than or equal to 0.243. While this result is, indeed, slightly less than 0.306 for 20 samples, it does tell us that 20 samples give a reasonably accurate answer. Indeed, we will need to tighten some of the sensitive tolerances after all.

The compensator is the back focus, and a total range of ± 0.23 mm was encountered with a standard deviation of 0.14 mm.

Surface Irregularities

In Chap. 4 we discussed the concept of optical path difference and its influence on image quality. One of the most important and influential rules of thumb is the Rayleigh criteria which tells us that if the peak-to-valley optical path difference is less than or equal to one-quarter wave, then the image quality will be nearly indistinguishable from perfect diffraction-limited performance.

Optical path difference (OPD) can be introduced by the following factors:

- The fundamental *aberrations present in the basic lens design* such as spherical aberration, coma, astigmatism, defocus, and other image-degrading aberrations.
- *Assembly and alignment errors*. Included in this category are the various tolerance forms such as element thicknesses, airspaces, wedges, tilts, decentrations, and other tolerance types.
- *Environmental effects* such as thermal soaks and/or gradients.
- *Effects external to the system* such as atmospheric turbulence.
- *Surface irregularities and other wavefront errors not included in the previous errors*. Included here are the residual manufacturing errors, which cause a surface to deviate or depart from its nominal shape. Most often, this is a deviation from sphericity or flatness, but it can, of course, also be a deviation from a prescribed aspheric profile if the nominal surface is aspheric.

The basic lens design performance residual, which is due to the various forms of aberrations, will consist of the orders of aberration present in the design such as defocus, spherical aberration, coma, astigmatism, field curvature, and the chromatic aberrations of axial and lateral color which are changes in the basic aberrations with wavelength. Thus, the design of a double gauss camera lens may have a mix of third, fifth, and higher orders of aberration, both on axis with spherical aberration, as well as off axis with other aberrations.

Recall that the OPD polynomial is one exponent higher than the transverse ray aberrations. In other words, third-order spherical aberration affects the wavefront proportional to the fourth power of the aperture, third-order astigmatism is linear with aperture, and the effect to the wavefront is quadratic with aperture, and so on. The net

result is that the basic lens design will have a mix of all of the residual aberrations present in the prescription. Fortunately, lens design software programs are quite robust and can model these aberrations and accurately predict the MTF and other measures of performance. The optical path difference introduced by our design prescription is accurate and quantifiable.

The optical path difference introduced by assembly and alignment errors include symmetrical errors such as power fit to testplates, element thicknesses, and airspaces, as well as asymmetrical errors such as element wedge, tilt, and decentration. In most cases, the effects of these tolerances will be the introduction of primarily low-order aberration and, to a reasonable extent, this will be true throughout the system. For example, the resulting effect of the designated number of fringes of the power fit to testplate will, to a very large degree, contribute to the on-axis image defocus and, to a lesser extent, third-order spherical aberration. The axial image will also show the effects of the asymmetrical tolerances, including wedge, tilt, and decentration. These perturbations will contribute primarily low-order coma and astigmatism to the axial image. Similarly, off axis we will see the introduction of mostly lower-order aberrations due to the manufacturing and alignment errors.

The results of environmental effects, such as thermal soaks and gradients, will also be primarily low order.

This brings us to *surface irregularities*, which we will divide into several categories:

- *Conventional optical manufacturing.* Here we are talking about spherically surfaced lenses in approximately the 10- to 30-mm-diameter range manufactured in a production environment using conventional machinery. For longer radii, elements are blocked for cost economies. The residual surface irregularities for these elements are primarily cylindrical in form. Thus, the surface is literally a toroid which is best thought of as a cylindrical departure from sphericity. Clearly, to the axial imagery, this will introduce astigmatism, and the aberration is of low order.
- *Larger elements.* As lens elements become larger, the surface irregularities tend to depart some from the classical cylindrical shape. We can have asymmetries, which can in many cases become more highly asymmetric and nondescript and less well correlated.

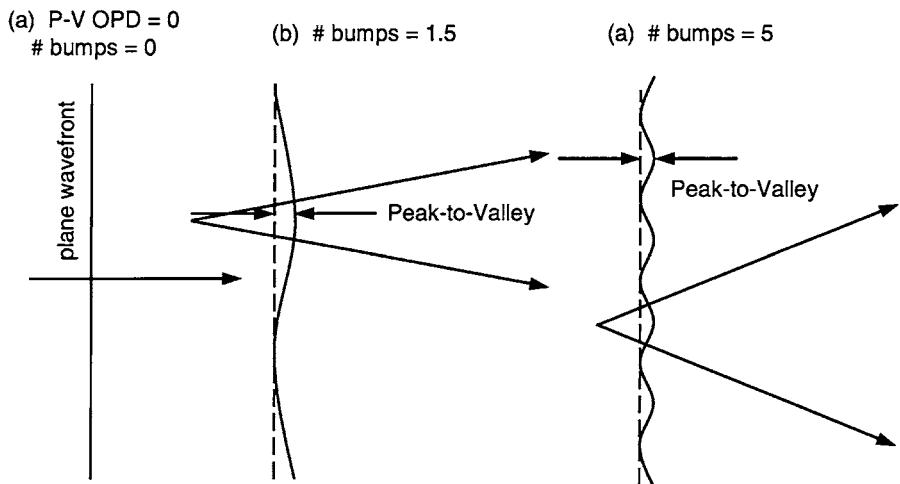
- *Larger surfaces such as telescope mirrors.* Whenever larger surfaces are involved, especially if there are aspheric surfaces, we often encounter much higher-order surface irregularities. These effects can, for example, be due to the manufacturing process where subdiameter polishing tools are often used. In the process of reaching the desired surface profile (often a paraboloid or similar conic section, sometimes with an intentional higher-order residual), high-frequency irregularities are often left in the surface.
- *Thin lenses, windows, mirrors, and plastic optics.* Very often, we require very thin lenses, windows, and mirrors, and also in this category we have injection-molded plastic lenses (as well as compression-molded glass lenses). We may, for example, use flat glass manufactured by a process called *float glass*. In many of these cases the residual surface irregularity is less straightforward to predict and can often have nonrotational symmetrical residuals from the nominal surface shape.

The primary difference between small, high-production lenses, where a cylindrical departure from sphericity results, and some of the latter examples, such as injection-molded plastic lenses and thin-float glass mirrors or beamsplitters, is that in these latter scenarios we may have both larger departures from the ideal surface profile as well as a lower “correlation” surface due to more “bumpiness” on the surfaces and therefore to the wavefront.

How Does Correlation Relate to Performance?

Consider Fig. 16.12 where we show three sinusoidal waveform models: Fig. 16.12a is plane or flat, Fig. 16.12b has a given peak-to-valley surface irregularity with 1.5 bumps across the surface, and Fig. 16.12c has the same P-V irregularity with five bumps across the surface. The term “correlation,” in very simple terms, is the inverse number of bumps across the surface. Thus, the surface in Fig. 16.12b has a correlation of 0.666 and the surface in Fig. 16.12c is less correlated with a correlation of 0.2.

Figure 16.12
Sinusoidal Wavefront Shapes



If the preceding represents the deviation of the wavefront at the exit pupil from its ideal spherical shape, then multiplying the total angular ray deviation by the focal length of our optical system will give the maximum image blur diameter or extent at the image. The larger the slope errors, the larger the image blur. Unfortunately, these effects are often random and change from part to part or system to system, so some form of modeling and approximation is in order.

Effect to Spot Diameter

Assume that we have a wavefront that departs from perfect flatness or sphericity sinusoidally, as shown in Fig. 16.12. Assume the following:

- A = the peak-to-valley variation of the wavefront
- n = number of bumps over the wavefront extent
- D = the total extent of the wavefront (exit pupil diameter)

The maximum slope of the wavefront can be shown to be

$$\text{Maximum slope} = \frac{\pi A \lambda n}{D}$$

Thus, the total angular spread of the rays proceeding to the image is twice the preceding result. If we assume that the rms wavefront error is one-fifth of the peak-to-valley (a reasonable assumption), the angular spot diameter containing 100% of the light is

$$100\% \text{ spot diameter (rad)} = \frac{5\pi\sigma\lambda n}{D}$$

where σ is the rms wavefront error. If we now assume that the energy is uniformly distributed in the image, we can rewrite the previous relationship in its final and most useful form:

$$\text{Spot diameter} = \sqrt{p} \left(\frac{5\pi\sigma\lambda n}{D} \right)$$

where p is the fraction of the total energy. This gives us the diameter of the spot inside of which there is a given percent energy.

Let's do a quick sanity check. Assume we have a wavefront with three bumps, which just meets the Rayleigh criteria exiting an $f/10$ lens with a 100-mm exit pupil diameter. The rms wavefront error is thus 0.05 wave, and the wavelength is 0.5 μm . Our focal length is 1000 mm. Thus, the diameter of the spot containing 80% of the energy is

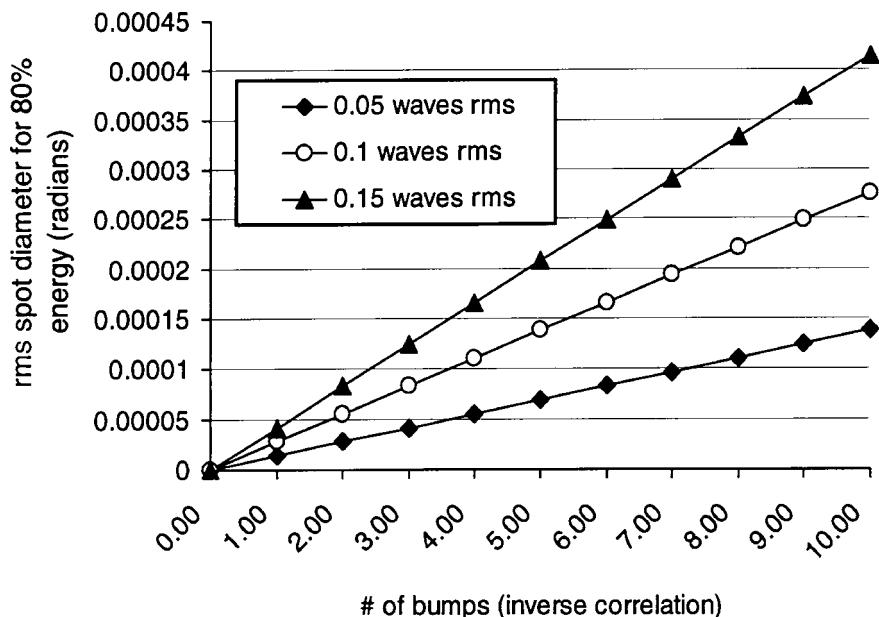
$$\begin{aligned} 80\% \text{ spot diameter} &= \sqrt{0.8} \left\{ \frac{5 \times 3.141 \times 0.05 \times 0.5 \times 3}{100,000} \right\} \times 1000 \\ &= 0.0105 \text{ mm} \end{aligned}$$

The diffraction-limited Airy disk diameter for the previous $f/10$ lens is approximately 0.012 mm, which is very close to the derived value of 0.0105 for a situation which should be essentially diffraction limited. As the number of bumps increases or the other parameters change, we can compute the approximate predicted blur diameter. While this derivation is not rigorous, and it is based solely on geometry, it is extremely useful when you really do not know the exact form of the wavefront error but you do have an idea of the correlation and the wavefront error.

Figure 16.13 shows graphically the 80% energy blur diameter, in radians, as a function of rms wavefront error and the number of bumps on the wavefront. These data are based on the derivation shown earlier.

Figure 16.13

Predicted Blur
Diameter As a
Function of rms
Wavefront Error and
Number of Bumps



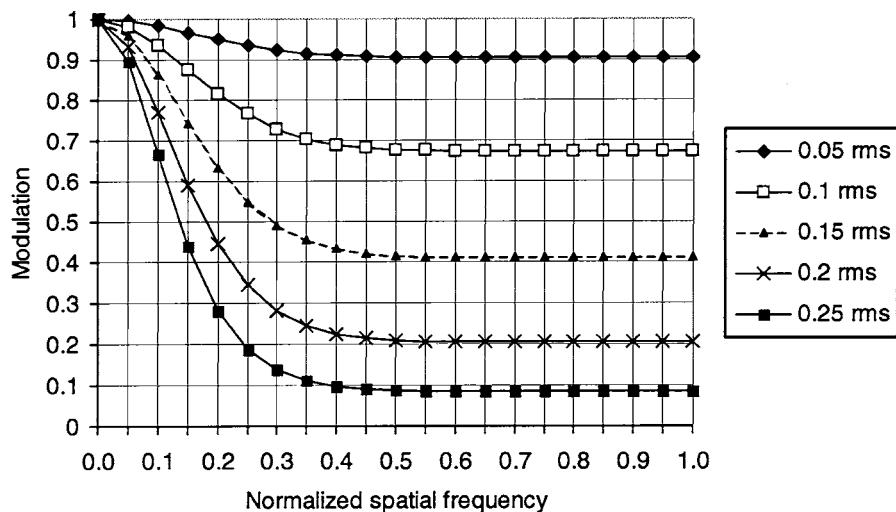
Results smaller than the Airy disk diffraction diameter are fictitious and in these situations the prediction should revert back to the Airy disk diameter.

Effect to MTF: The Optical Quality Factor

A number of years ago, Hufnagle of Perkin-Elmer developed an empirical relationship whereby the MTF degradation can be derived as a function of the rms wavefront error and the correlation of the wavefront. This relationship yields what has become known as the optical quality factor (OQF), and is given by

$$\text{OQF} = e^{-(2\pi r_{\text{rms}})^2 (1 - e^{-2n^2 s^2})}$$

Figure 16.14
OQF As A Function of
rms Wavefront Error
and Number of
Bumps



where n is the number of bumps over the exit pupil (the inverse correlation) and s is the normalized spatial frequency relative to cutoff. Figure 16.14 shows the OQF as a function of normalized spatial frequency and rms wavefront error for a correlation of 0.333, or three bumps across the pupil. If we multiply these data by the perfect system MTF we get the results in Fig. 16.15.

Figure 16.15
Predicted MTF As
Function of rms
Wavefront Error and
Number of Bumps

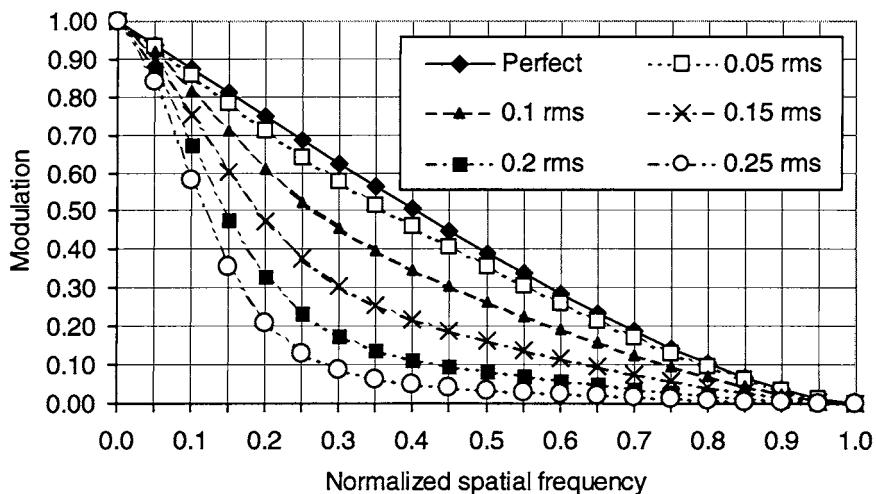
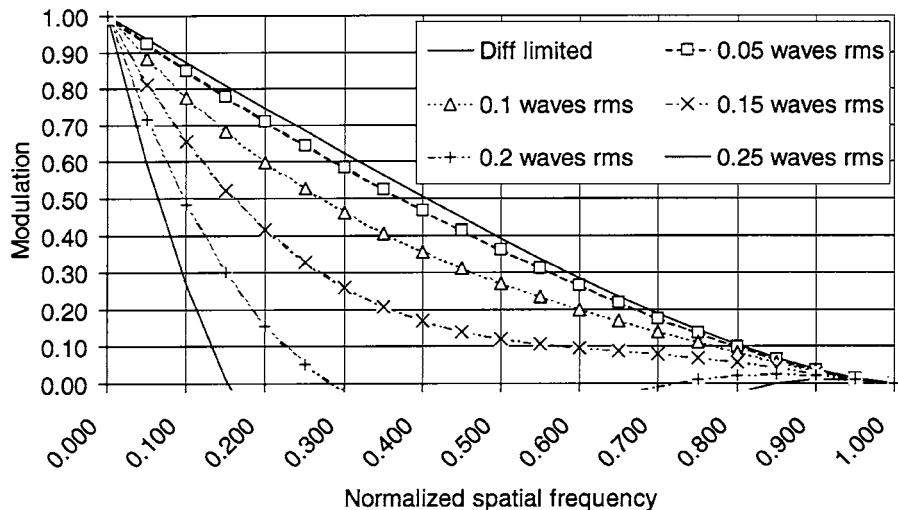


Figure 16.16
MTF Versus rms
Wavefront Error



There is an empirically derived approximation by Shannon which can also be used, and this is shown in Fig. 16.16. The equation for this data is

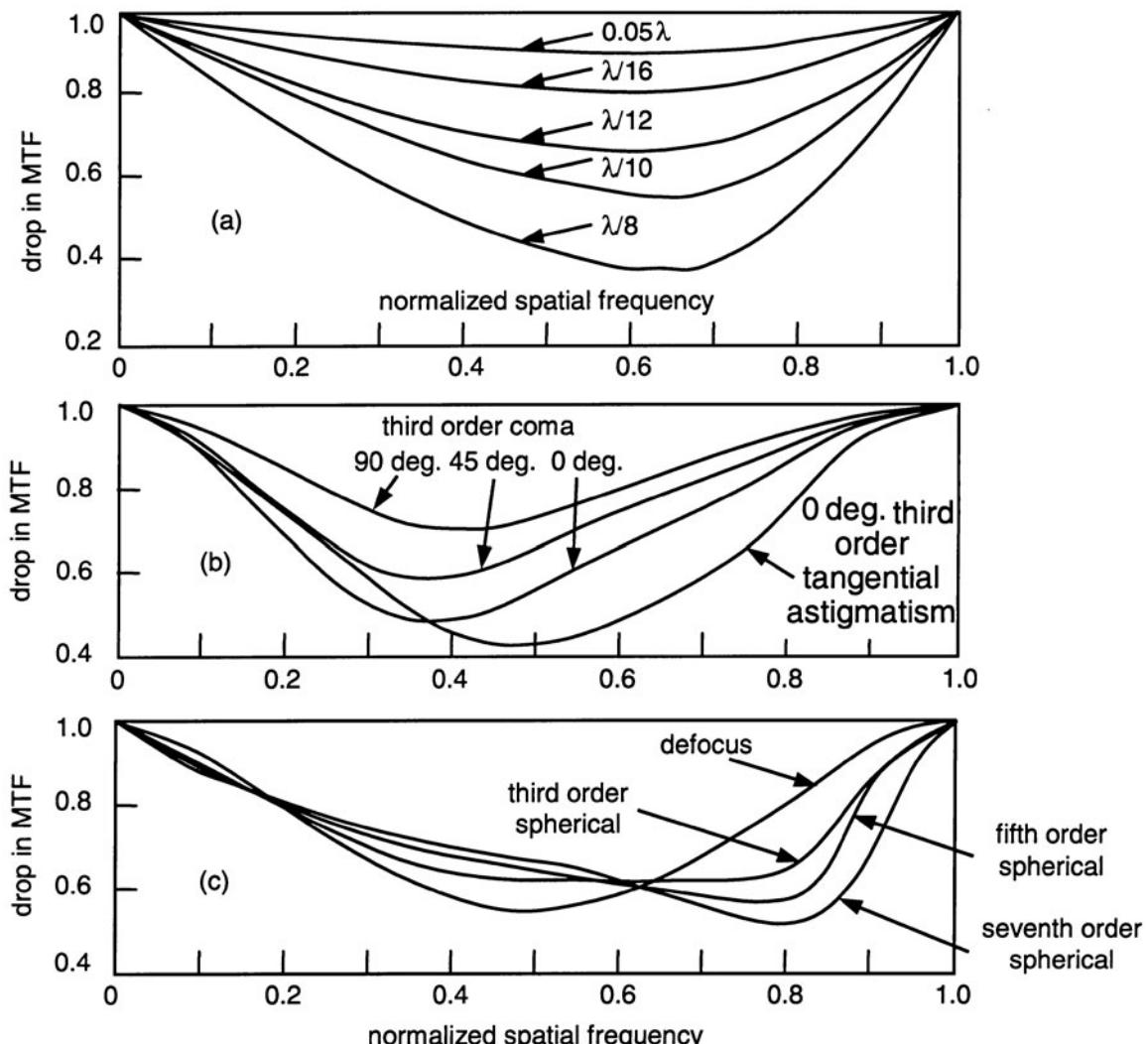
$$\text{MTF}(\nu) = \frac{2}{\pi} \left[\arccos(\nu) - \nu \sqrt{1 - \nu^2} \right]$$

$$\text{ATF}(\nu) = \left\{ 1 - \left[\left(\frac{W_{\text{rms}}}{0.18} \right)^2 \right] [1 - 4(\nu - 0.5)^2] \right\}$$

$$\nu = \frac{\text{spatial frequency}}{\text{cutoff frequency}} = \frac{N}{[1/(\lambda) f/\#]}$$

In this relationship, the diffraction-limited MTF is given by the first equation as a function of the normalized spatial frequency, ν . The equivalent to the OQF described earlier is given by the second equation as a function of the normalized spatial frequency ν and the rms wavefront error.

Another form of MTF degradation, or OQE, is shown in Fig. 16.17a, where we show the effect of a mix of third-, fifth-, and seventh-order aberrations. The wavefront degradations are in the form of rms wavefront error. In Fig. 16.17b and c we show the OQE for 0.1 wave rms of nonrotationally symmetric aberrations and rotationally symmetric

**Figure 16.17**

MTF Drop As a Function of Wavefront Error

aberrations, respectively. It is interesting to note that for 0.1 wave rms the MTF drops to approximately 0.6 of its nominal value, regardless of the form of aberration.

The data in this section can be very useful for predicting the MTF drop due to wavefront errors which you may have derived from a

performance error budget, but you may not know the specific form of the error so you cannot model it directly on your computer program. For example, if you were to have approximately one-half wave P-V of random irregularity due to a mirror in your system, this equates to approximately 0.1 wave rms, and the OQF at midfrequency would be in the order of 0.6. These data allow us to quickly and easily assess and predict the effect of wavefront errors on the optical system.

Beam Diameter and Surface Irregularity

There is one additional important point, which needs to be discussed—how the beam diameter relates to the surface and wavefront irregularity. The critical factor is how much wavefront error is introduced to the wavefront as it proceeds through our system, surface by surface. A parameter that becomes extremely important is the beam diameter or the “footprint” of light at each surface. Consider the lens in Fig. 16.18. Note how almost the entire aperture of the cemented doublet is used regardless of the field position. The element closest to the image is nearly the same diameter as the doublet; however, the beam diameter going to any given image point uses only about 20% of its diameter.

The impact of this is clear from Fig. 16.19a, where we show a simulated interferogram with five waves of astigmatism. Let us assume that an equivalent wavefront error is introduced by the final lens element over its full diameter. Since the beam diameter reaching any point in the field of view is only 20% of the lens diameter, what we have is the

Figure 16.18
Beam Diameter

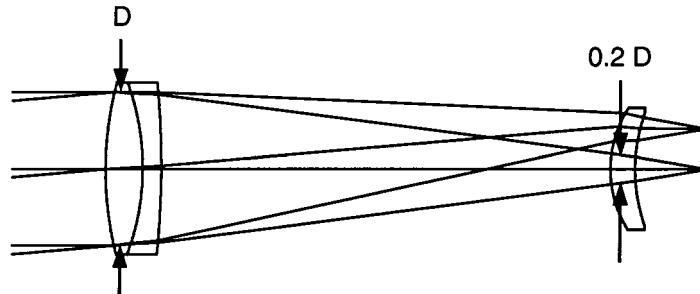
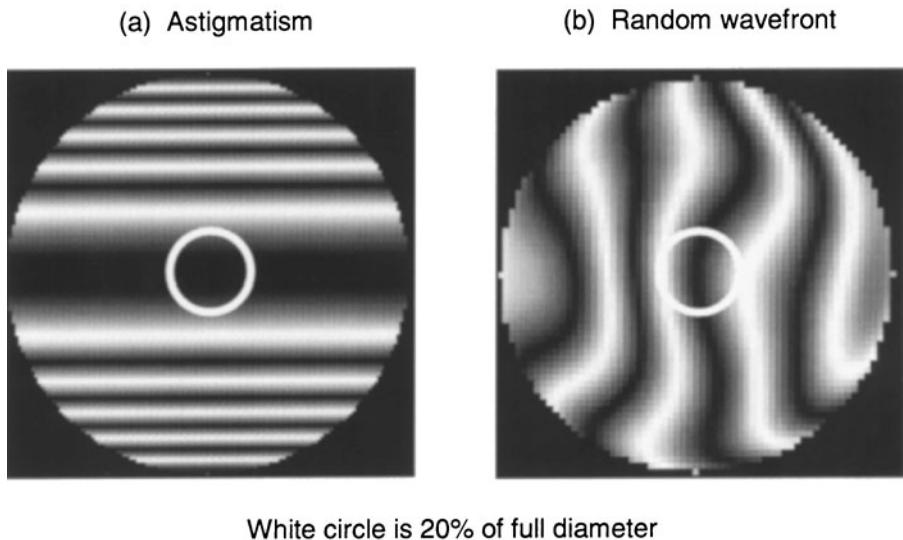


Figure 16.19
Reduction of OPD
for Smaller Beam
Diameters



effect of the same interferogram but only within the white circle. If we have astigmatism, as shown, the net wavefront error to the system will be reduced from 5 waves to 0.2 wave since the wavefront polynomial for astigmatism is quadratic with aperture, and $0.2^2 = 0.04$, which means that only one-twenty-fifth of the astigmatism is introduced to the wavefront.

Figure 16.19b shows a random wavefront where we can also see that, over a reduced beam footprint, the net wavefront error can be significantly reduced. In this case, the full diameter has a residual of approximately 0.8λ peak-to-valley, and over 20% of its aperture (the white circle) we find approximately 0.02 wave, a reduction of 40 times.

The Final Results

Once we predict the net system degradation, the performance can be shown in various formats. In order to show how these results all compare, consider Figs. 16.20 and 16.21, where we show the performance for a perfect $f/5$ system with no diffraction or aberrations (a), as well as a diffraction-limited system (b), 0.25λ P-V (c), 0.5λ P-V (d), and 0.75λ (e).

Figure 16.20

Image Quality for Different Amounts of Peak-to-Valley Third-Order Spherical Aberration

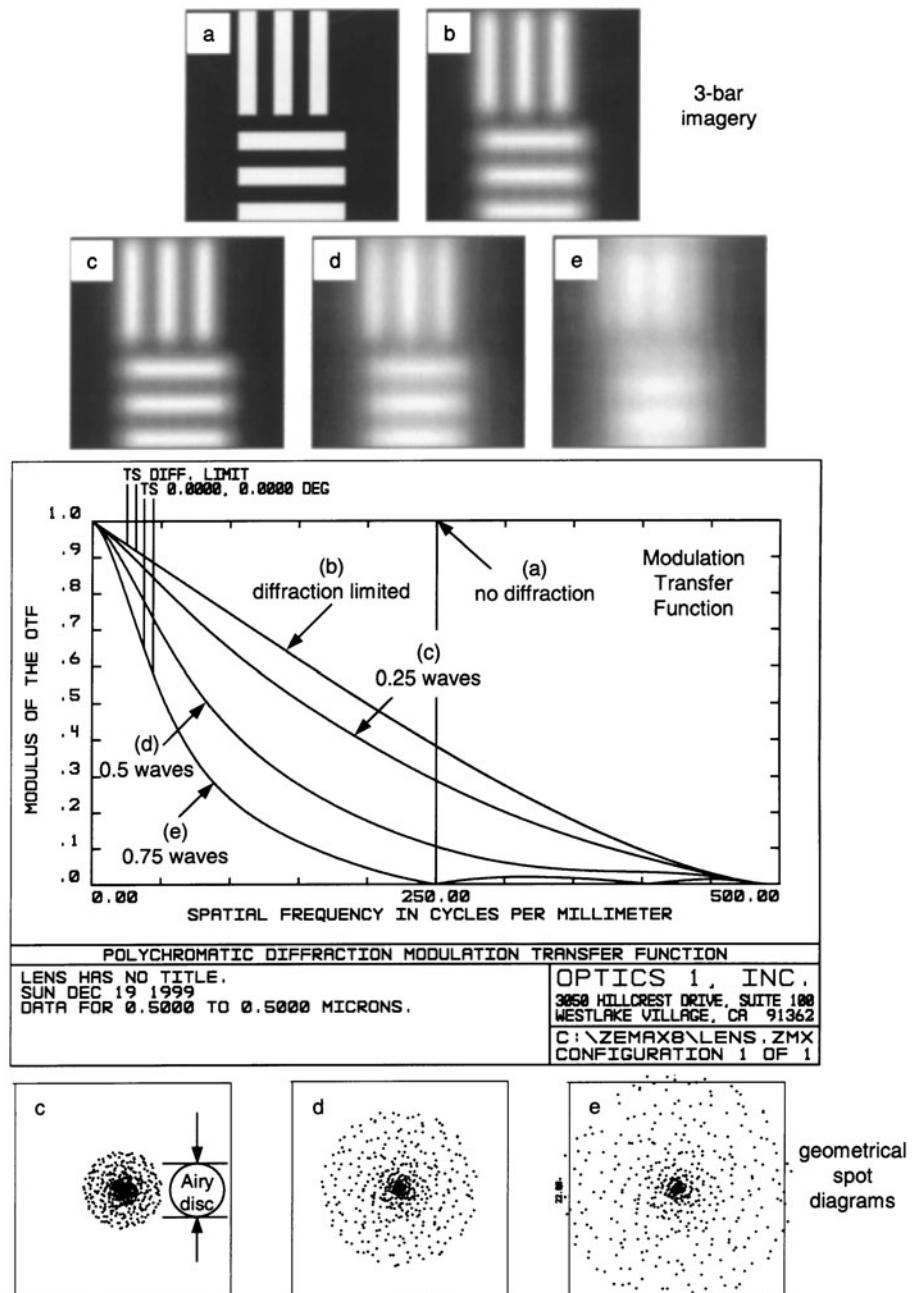
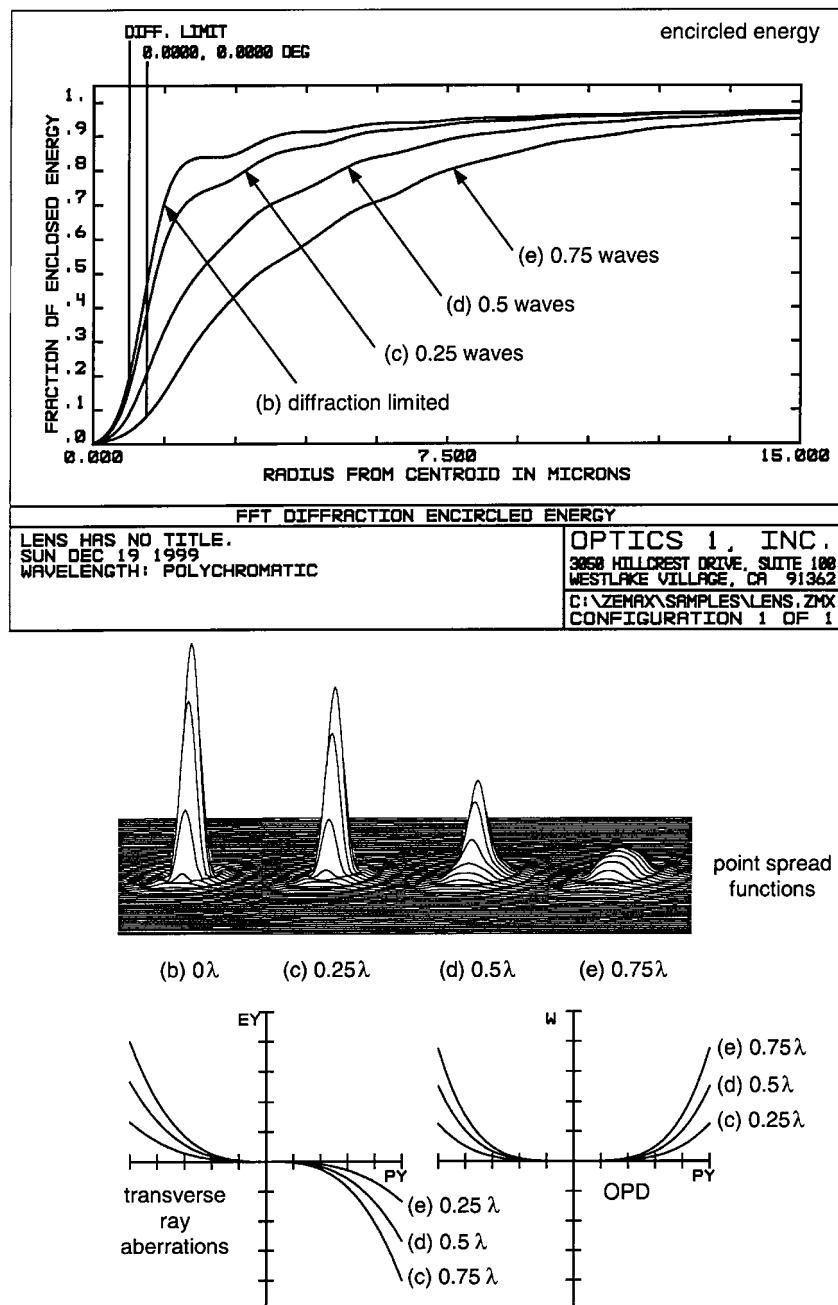


Figure 16.21
Image Quality for
Different Amounts of
Peak-to-Valley Third-
Order Spherical
Aberration



The aberrations are all third-order spherical elements. The following performance metrics are shown in Figs. 16.20 abd 16.21:

- *A graphical representation of a three-bar Air Force–type target whose bar pattern is selected to be at 50% of the cutoff spatial frequency at 250 line pairs/mm.* Note that for these data, as well as the other data presented here, the 0.25-wave situation is only slightly degraded from the diffraction-limited metric, as predicted by the Rayleigh criteria. It is interesting to note that if there were no diffraction effects, and we had an aberration-free system, then the three-bar pattern would image as shown in Fig. 16.20a. This corresponds to a modulation of unity and, for all practical purposes, represents the target for the data in Figs. 16.20b through d.
- *A plot of the MTF.* The reduction in contrast is quite evident, reaching zero for the 0.75-wave case. Note how the bar pattern imagery correlates with the MTF data.
- *Geometrical spot diagrams, with the Airy disk size shown for reference.* Note in the case of the spherical aberration that we have here somewhat of an intense center to the pattern rather than a uniform intensity pattern which would be evident from pure defocus. As the pattern grows from the Airy disk diffraction pattern diameter, the MTF drops and the contrast degrades, as already shown.
- *A plot of encircled energy.* Here, too, we can see the increasing blur diameters as the spherical aberration increases. As before, the 0.25-wave case is close to diffraction limited, as predicted by Rayleigh.
- *Plots of the point spread function (PSF).* This is one of the more instructive examples of the Rayleigh criteria. The PSF for the perfect system and the 0.25-wave system are very nearly identical in overall appearance. While we do see a drop of intensity at the center of the pattern, the overall appearance of the PSF is nearly the same as for the perfect system. The first ring is still quite evident. However, as soon as we reach 0.5 wave and above, the whole character and appearance of the pattern degrades, and more energy is thrown out from the central maximum of the pattern.

- *Plots of the transverse ray aberrations.* Here we show the transverse ray aberrations for the different amounts of wavefront error. These data are cubic with aperture.
- *Plots of the optical path difference.* Finally, we show the optical path difference for the different amounts of wavefront error. These data are proportional to the fourth power of the aperture.

All of the data here are for *precisely* the same system with the same third-order spherical aberration. These data, along with the OQF and other relationships discussed earlier in this chapter, can be used to help predict the performance of imaging optical systems for various amounts of wavefront error.

Optomechanical Design

An optical system cannot function without structure and related mechanical parts to hold the optics together in proper form and alignment and to provide any mechanisms needed for operation as an optical instrument. Through optomechanical design, we integrate and package the optical and mechanical systems into a synergistic ensemble that will accomplish some stated purpose in a defined environment. It is this optomechanical aspect of optical systems design that is addressed in this chapter.

Environmental Considerations

Most optical instruments are expected to survive, without any damage, certain extreme environmental conditions and to perform fully to specifications in other, less stringent, operational environments. It is important to understand the mode of transportation envisioned for the instrument because shipping usually entails exposure to adverse environmental conditions—especially temperature, vibration, and shock. We here consider briefly how various important environmental conditions impact the instrument design.

TEMPERATURE The temperature of an optical instrument depends largely upon that of its surroundings. Heat transfers into or out of the

instrument's components by conduction through mechanical contacts, convection within the surrounding air, or radiation between bodies at different temperatures. Heat-producing components such as light sources, motors, electronics, and radiation from external sources such as the sun also may affect instrument temperature.

Temperature extremes encountered by optical instruments on or near the Earth's surface generally range from about -62° to 71°C . If human operation is involved, the temperatures usually are limited to -54° to 52°C . Specifications frequently reflect these temperature ranges for the storage and operational use, respectively, of optical systems. Equipment intended for use in more benign environments should be designed for the conditions of shipment that may involve temperature extremes of at least -32° to 52°C . In space, optics may be required to function at temperatures approaching absolute zero at -273.1°C . They also might be subjected to variations of radiant heat input from the sun, as in the case of a sensor orbiting the Earth and moving into and out of the Earth's shadow.

Many things happen within an optical assembly when its temperature changes. Changes occur in optical surface radii, optical component diameters, thicknesses of optical components, refractive indices of refracting materials, refractive indices of the surrounding media (usually air), dimensions of mechanical parts, and sizes of air spaces. Each change from the temperature at assembly tends to affect focus and alignment of the optics so performance may suffer.

In many cases, the temperature of an instrument never reaches equality with that of its environment nor uniformity within itself. Spatial and/or temporal gradients therefore may occur and these may decenter or tilt the optics, cause pointing errors, or distort optical surfaces. Temperature changes also modify the mounting forces applied to optical parts at the time of assembly. Extremely rapid changes in temperature can permanently damage the instrument.

PRESSURE The major effects of pressure on optics are deformations of optics exposed to pressure differentials, variations of refractive index of air with altitude, and contamination by moisture and/or dirt forced by pressure differentials through minute leaks in the instrument's enclosure. Pressure exposure for optics can range from a vacuum to >1000 atmospheres, depending upon the application.

MOISTURE, CONTAMINATION, AND CORROSION The surfaces of many optical coatings, some glasses, most optical crystals and some

metals can be degraded by interaction with atmospheric moisture. This moisture is quantified as relative humidity; it can range from 0 to 100%. An extreme case is immersion of the instrument to a specified depth (and hence to a corresponding pressure) under water. Condensed water droplets as well as particulates such as dust and chemical deposits on optical surfaces tend to absorb light and increase light scatter, thereby reducing both image contrast and light transmission.

Contaminating foreign matter of molecular nature may result from outgassing of materials. This is especially critical for space borne equipment. As a rule of thumb, all materials used in space should have <1% total mass loss and 0.1% collected volatile condensable material when tested per ASTM E595-93.¹ Guidance on general contamination issues may be found in Tribble.²

Metallic materials in contact should be compatible with regard to their tendency to form a galvanic couple in the presence of moisture. If mutually active materials must be contacted, the interface should be protected by a suitable intermediate finish or coating.

Effective sealing will help reduce internal effects of moisture and other contaminants. Frequently, instruments are purged after assembly with dry gas such as conditioned air, N₂ or He. In some cases, the internal pressure may be increased over the external ambient pressure. This practice is of questionable value since water vapor tends to diffuse through seals and housing walls in proportion to the partial pressure differential of the internal and external atmospheres.

ABRASION AND EROSION Optical coatings and surfaces exposed to wind-driven dust, sand, rain, or ice crystals may be damaged. This most frequently occurs for windows or lenses on optical devices mounted externally on ground vehicles, helicopters, and other aircraft. Most crystalline materials used in infrared sensors are especially vulnerable. Removable mechanical covers can afford protection when the instrument is not operating. Some degree of protection for vulnerable optics results from special surface coatings.

In the space environment, micrometeorites, atomic oxygen, and debris can also damage optics. Durable optical coatings and protective shields will help avoid this damage.

FUNGUS Optics exposed to fungus spores in a hot, high-humidity environment can be damaged by growths on their surfaces. Organic materials such as cork, leather, and natural rubber should never be used

in or on optical instruments because they are especially susceptible to fungus growth. Fingerprints and other surface contaminants also may serve as nutrients for fungus. Periodic cleaning of all exposed surfaces will help prevent damage from fungus. Care must be exercised in cleaning optics.

HIGH-ENERGY RADIATION The transparencies of many refracting materials are reduced when they are exposed to high-energy radiation such as gamma and x-rays, neutrons, protons, and electrons. Limited protection can be provided for such optics by shielding them with materials that absorb the radiation. A few types of radiation-protected optical glasses containing cerium oxide are available. These glass types darken minimally when exposed to radiation. Mechanically, these protected glasses are essentially the same as their conventional counterparts.

Energetic laser beams can damage optics and coatings, especially if the surfaces are not clean or if they have imperfections such as scratches or pits. Molecular absorption of laser radiation within materials can cause transmission losses. For example, exposure to intense UV laser light of a fused silica lens containing oxygen as an impurity can cause progressive darkening that curtails usefulness of that optic.

VIBRATION AND SHOCK Mechanical forces imposed upon an optical instrument in cyclic manner (vibration) or as short pulses (shock) can displace optics and other components elastically or permanently disturb their alignment. In extreme cases, acceleration forces can cause catastrophic damage.

A design characteristic of importance here is the structural stiffness of the optomechanical system and its components. A stiff structure has a high natural, or fundamental, frequency so is less likely to resonate and incur misalignment or damage when driven externally.

A shock may be defined as a force loading of duration equal to or less than the period (the reciprocal of the natural frequency) of the assembly. In general, an optical instrument should be designed to withstand shock accelerations of at least 15 times ambient gravity (g) in any of three orthogonal directions because such loading is common during shipping. Severe conditions for military and aerospace applications may be 100 to 500 times g while extremes can reach $\sim 11,000$ times g . Analytical techniques such as finite element analysis provide the best means for predicting and then evaluating the suitability of a given design in the specified vibration and shock environments. Analysis should always be verified by testing of the hardware.

Applicable Design Guidelines

For nonspace borne applications, general information as to extreme and typical environments to ~80 km altitude can be derived from the U.S. military specification MIL-STD-210³ or ISO Specification 10109.⁴ Either of these documents can serve as guides for the design of commercial or consumer optical equipment. For space applications, guidance can be derived from sources such as Tribble,⁵ Sarafin,⁶ and Shipley.⁷

Environmental Testing Methods

Testing is needed to confirm that an instrument will indeed meet specifications for resistance to environmental exposure. Detailed guidelines for testing may be found in US. military specification MIL-STD-810⁸ and in ISO specification 9022.⁹ Early knowledge of how an assembly or instrument is to be tested will often help guide the design of that hardware.

Mechanical Parameters and Properties

Following are definitions of important mechanical parameters and properties for the materials used in optical instruments. They are listed alphabetically and not necessarily in order of importance. Symbols used in equations in this chapter and their typical units are also included. Subscripts are added as appropriate to designate different applications.

BIREFRINGENCE measures inhomogeneity of refractive index within a refracting medium. It may occur naturally (in crystals) or may be caused by stress developed within isotropic materials. It is usually quantified as the optical path difference (OPD) between the orthogonal states of transmitted polarized light in nm per cm of path length.

DENSITY (ρ) is mass per unit volume. It is expressed in g/cm³.

FORCE (F) and PRELOAD (P) both represent an external influence exerted on a body that results primarily in acceleration or deformation of shape. It is expressed in newtons (N).

POISSON'S RATIO (ν) is the ratio of lateral unit strain to longitudinal unit strain in an object under uniform tension or compression. It is dimensionless.

STRAIN ($\delta L/L$) is the dimensional change per unit length of an object induced by externally applied force. It is dimensionless.

STRESS (S) is the force per unit area imposed on an object or generated internally, as in the case of poorly annealed glass. Units are pascals (Pa), which are equivalent to N/m².

THERMAL CONDUCTIVITY (κ) is the quantity of heat transmitted per unit of time through a unit area and unit material thickness per unit temperature gradient. It is expressed as W/(m-K).

THERMAL DIFFUSIVITY (D) is a measure of the rate of heat spread through a material. It equals thermal conductivity divided by the product of density and heat capacity. Units are m²/s.

THERMAL EXPANSION COEFFICIENT (α) is the change in an object's length per unit length per degree of temperature change. Typical units are mm/mm per °C (or ppm/°C).

YIELD STRESS (S_y) is the value of internal stress, in Pa, for a material at which permanent deformation (strain) of 0.2% dimensional offset occurs.

YOUNG'S MODULUS (E) is the rate of change of unit tensile or compressive stress with respect to linear strain within the elastic limit. It is a measure of mechanical stiffness of a body. Units are Pa.

Typical Mechanical Property Values for Selected Materials

Tables 17.1 and 17.2 list typical values for some mechanical properties of selected common refractive and mechanical materials that are used in optical instruments. Table 17.3 gives mechanical properties and values of certain factors of merit (FOM) for materials most frequently used in mirror

TABLE 17.1

Key Mechanical Properties of Selected Refractive Materials

Material	Young's Modulus	Poisson's Ratio	CTE	Density
Units	MPa	—	mm/(mm·°C)	g/cm ³
NFK5 glass	6.2E+4	0.232	9.2E-6	2.45 L
K10 glass	6.5E+4	0.190 L	6.5E-6	2.52
NZK7 glass	7.0E+4	0.214	4.5E-6 L	2.49
NBK7 glass	8.2E+4	0.206	7.1E-6	2.51
SF2 glass	5.5E+4 L	0.227	8.4E-6	3.86
NLaF2 glass	9.4E+4	0.288 H	8.1E-6	4.30
NLaF7 glass	9.6E+4 H	0.271	7.3E-6	3.73
NSF4 glass	9.0E+4	0.256	9.5E-6 H	3.15
LaSFN9 glass	1.1E+5 H	0.286	7.4E-6	4.44
Fused silica	7.3E+4	0.164	0.52E-6	2.20
AMTIR 1	2.2E+4	0.266	12.0E-6	4.4
AsS ₃	1.58E+4	0.295	26.1E-6	3.43
BaF	5.32E+4	0.343	18.4E-6	4.89
CaF ₂ *	7.6E+4	0.260	9.2E-6	3.18
Ge	1.04E+5	0.278	6.0E-6	5.32
MgF ₂	16.9E+4	0.269	14.0/8.9**	3.18
Sapphire	4.0E+5	0.27	5.3E-6	3.97
Si	1.31E+5	0.279	2.6E-6	2.33
ZnS	7.45E+4	0.29	4.6E-6	4.08
ZnSe	7.03E+4	0.28	7.1E-6	5.27

*Values are for Schott *Lithotec-CaF₂* developed especially for use in UV lithographic optics.

**Birefringent material. Values are for orthogonal directions in crystal.

substrates. Glasses are Schott varieties. Except for the ubiquitous NBK7, they were selected for having the lowest (L) or highest (H) value for the indicated property from a list of 49 commonly used optical glasses in Yoder.¹⁰ These tables can be used to advantage during optical system design to compare candidate material characteristics and to help in making

TABLE 17.2

Key Mechanical Properties of Metallic and Composite Structural Materials Frequently used in Optical Instruments

Material	Young's Modulus MPa	Poisson's Ratio —	CTE mm/(mm°C)	Density g/cm ³	Specific Stiffness* (FOM) —
Aluminum 6061-T6	6.82E+4	0.332	2.36E-5	2.68	1.00
Aluminum 2024	7.31E+4	0.33	2.29E-5	2.77	1.07
Beryllium I-70H	2.89E+5	0.08	1.13E-5	1.85	6.32
Beryllium copper	1.27E+5	0.35	1.78E-5	8.25	0.62
Brass (typ.)	9.65E+4	—	2.05E-5	8.50	0.46
Graphite epoxy (typ.)	9.30E+4	—	2.00E-8	1.78	2.12
Invar 36	1.41E+5	0.259	1.26E-6	8.05	0.71
Magnesium MIA	4.48E+4	—	2.52E-5	1.77	1.02
Molybdenum TZM	3.18E+5	0.32	5.00E-6	10.2	1.26
Stainless steel 304	1.93E+5	0.27	1.47E-5	8.00	0.98
Stainless steel 416	2.00E+5	0.283	9.90E-6	7.80	1.04
Titanium 6Al4V	1.14E+5	0.34	8.80E-6	4.43	1.04
CESIC	2.35E+5	—	2.60E-6	2.65	3.59
SXA metal matrix (typ.)	1.17E+5	—	1.24E-5	2.90	1.63

*Equals E/ρ. Values are relative to Al 6061.

appropriate material choices. Additional parameters for the listed materials and values for other materials are given by Yoder¹⁰ and Paquin.¹¹

Structural Design

Any structure can be thought of as an assemblage of basic elements such as plates, shells, tubes, rings, and beams. The structure may be created by casting, by machining from a single billet of material (commonly called hogging-out), or by connecting individual parts. An integral structure is generally stiffer than the equivalent one built up from many parts. Joints are progressively weaker if accomplished by welding,

TABLE 17.3

Mechanical Properties and Values of Merit Factors (FOM) for Materials Commonly used in Mirror Substrates

Property:	Young's	Poisson's	CTE	Density	FOM	FOM	FOM
	Modulus	Ratio			Steady State*	Transient*	Specific Stiffness
	Units: MPa	—	ppm/°C	g/cm ³	= α/κ	= α/D	= E/ρ
Desired Value:					Small	Small	Large
Fused silica	7.30E+4	0.17	5.8E-7	2.205	0.36	0.59	0.33
ULE 7971	6.76E+4	0.17	1.5E-8	2.205	0.02	0.04	0.30
Zerodur	9.06E+4	0.24	5.0E-8	2.53	0.03	0.07	0.36
Be I-70H	2.89E+5	0.08	1.13E-5	1.85	0.05	0.20	1.56
Al 6061	6.82E+4	0.332	2.36E-5	2.68	0.13	0.33	0.25
OFHC Cu**	1.17E+5	0.35	1.67E-5	8.94	0.53	0.14	0.13
Silicon	1.31E+5	0.42	2.6E-6	2.33	0.02	0.03	0.56
SiC***	3.3E+5	0.24	2.5E-6	2.89	0.01	0.03	1.14

*Refers to rate of temperature change

**Oxygen free high conductivity

***Reaction bonded, 30% Si

rivets, bolts, screws, or adhesive joints. A weaker structure will be affected to a greater degree by imposed accelerations (vibration and shock). Joined structures also provide imperfect heat transfer paths from one part to another, thereby increasing the likelihood for spatial temperature gradients to develop.

The mechanical structure of the optical instrument is subject to distortion due to gravity, acceleration, motions of internal parts, and thermal effects. These distortions may affect optical alignment and/or focus and thereby degrade instrument performance.

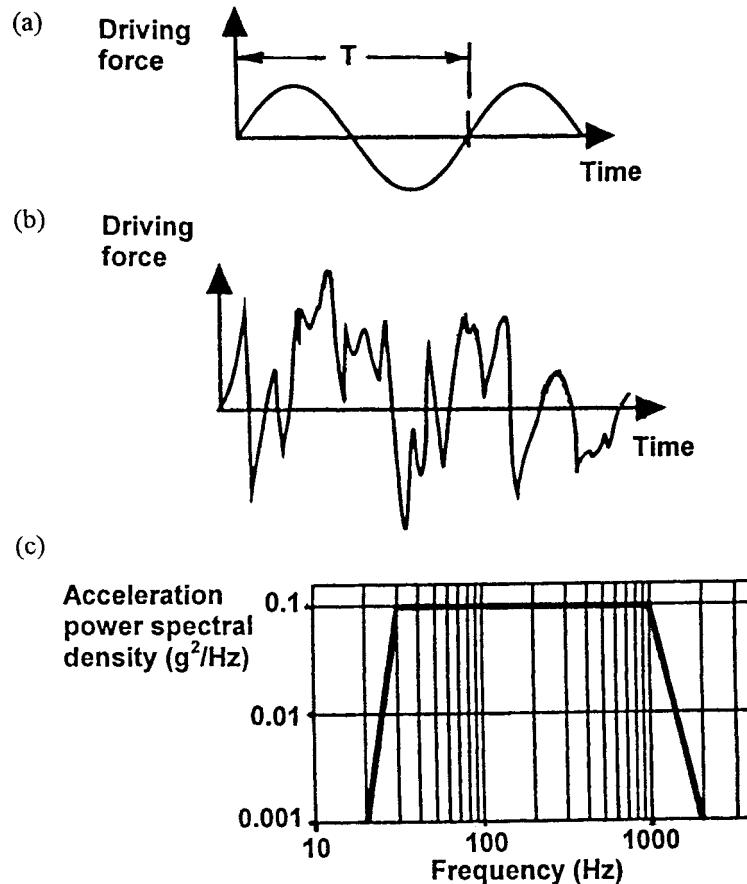
The weight of the structure frequently must be minimized. This may entail the use of thin wall sections, judicious placement of reinforcing webs, through holes or blind recesses machined into walls, or application of reduced density materials. One measure of the suitability of a material for a given structural application is its specific stiffness, which equals Young's modulus divided by density (that is, E/ρ). This factor should be as large as possible. See Table 17.2 for typical values for popular structural materials.

Vibration, Self-Weight Deflection, and Fundamental Frequency

Cyclic vibration is displacement of a body occurring periodically with a temporal period T such as represented by the sinusoidal curve of Fig. 17.1a. The first derivative of this curve is the body's velocity while its second derivative is the body's acceleration. The velocity and acceleration are also sinusoidal and have the same period, but are 90° and 180° respectively out of phase with the displacement. Any effect, such as friction, that resists motion is called *damping*.

An important condition occurs when the frequency of the driving force approximates the natural or fundamental frequency f_N of the body

Figure 17.1
Environmental Vibrational Inputs: (a) Periodic (Sine Wave)
(b) Random Vibration
(c) Acceleration PSD



acting as a structural system. Unless effectively damped, resonance then occurs and the body's vibratory amplitude will exceed that resulting from the driving force alone. The success of an instrument's design depends largely upon the designer's ability to predict and avoid these mechanical resonances. Typically, if the driving frequency differs from the natural frequency by at least a factor of two, amplification does not occur.

The following equation relates the system's approximate f_N to its mass m and its structural stiffness k :

$$f_N = (0.5/\pi)(k/m)^{1/2} \quad (17.1)$$

Here, k is in N/m, and m is in kg.

A stiff system of given mass has a large k and thus a high fundamental frequency. This is desirable because f_N may then significantly exceed the frequency of external vibrational disturbances from the environment or from moving mechanisms within the system. In that case, mechanical resonance will not be excited. In some cases where the excitation is only of high frequency, the fundamental frequency might well be made significantly lower than that driving frequency.

For example, if a video camera with a mass of 2.0 kg is supported in a mount so as to have a system k of 1.50E+5 N/m, its fundamental frequency would be $f_N = (0.5/\pi)[(1.50E+5)/2.0]^{1/2} = 43.6$ Hz. Applying the rule mentioned earlier, the system should not resonate from an external periodic force with frequency <21.8 Hz or >87.2 Hz.

The vibration environment may be random in nature instead of periodic. This means that, within a given range of frequencies, acceleration of some magnitude occurs at each frequency. See Fig. 17.1b. If the fundamental frequency of the instrument and its structure falls within this range, resonance could be excited.

Random vibrations are frequently quantified by their acceleration power spectral density (PSD). This is expressed graphically on log-log coordinates as a function that rises from zero, levels off, and falls again to zero at high frequencies. See the simple example in Fig. 17.1c. In more complex cases, different functions occur in different frequency regions. Over any selected range of frequencies, the acceleration PSD is quantified in units of g^2/Hz where g is a multiple of ambient gravity.

For a body vibrating randomly in a single degree of freedom, its rms acceleration response ξ can be approximated by the expression:

$$\xi = [\pi f_n \text{ PSD}/(4\eta)]^{1/2} \quad (17.2)$$

where PSD is defined over a specific frequency range and η is a factor quantifying the effective damping within that range. To illustrate use of this equation, let us assume that the camera and mount considered just above has a system damping factor of 0.05. As before, its f_N is 43.6 Hz. If the hardware is to be used in the random vibration environment indicated by Fig. 17.1c in the range 30 to 1000 Hz, the PSD is 0.1. From Eq. 17.2, $\xi = \{(\pi)(43.6)(0.1)/[(4)(0.05)]\}^{1/2} = 8.3$ times g. Vukobratovich¹² indicated that most structural effects result from the 3-sigma acceleration experienced. This camera/mount subassembly should then be designed and tested for accelerations of $(3)(8.3) = 25$ times g.

Shock

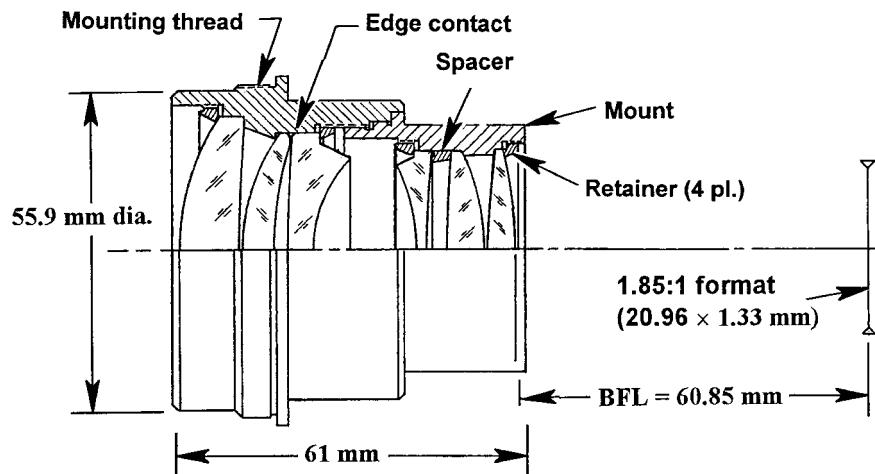
A force applied to a body suddenly and briefly is shock. It is expressed in multiples of ambient gravity and usually is an acceleration pulse lasting a small fraction of a second. It can be approximated by a half-cycle sinusoid. Resonance may be excited. Damage frequently results from severe impact such as might occur when an optical instrument falls to the floor. Shock transmitted through external structure to which the instrument is attached is less likely to do damage because the force is attenuated with distance and with the number of mechanically fastened joints encountered. The specification for a given instrument should indicate the level of acceleration at the instrument's mounting interface.

Rigid Housing Configurations

The traditional configuration for a small- to midsized optical instrument has the optics mounted individually and directly into a rigid (usually metal) housing. Adjustments needed for alignment are incorporated into the glass-to-metal interfaces. Intermediate cells may be used to facilitate assembly or adjustment. Although we term such a housing as rigid, it really is flexible on a microscopic scale. Walls and other structural members deform to some small degree under gravitational and other externally applied forces. Such deformations are acceptable if they do not cause rigid-body displacements or tilts of critical components in the optical system or deformations of those components exceeding applicable tolerances.

Figure 17.2

Sectional View of a Projection Lens Assembly Exemplifying Rigid Housing Construction (Courtesy of Schneider Optics, Inc., Hauppauge, NY.)



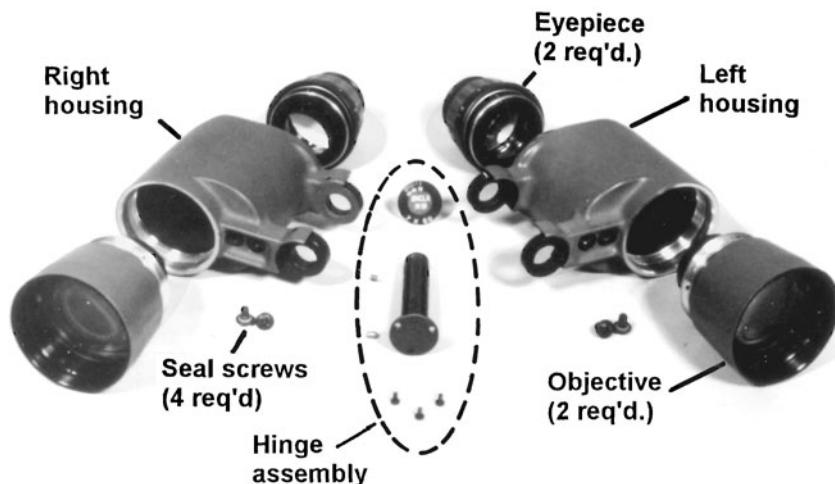
An example of a lens system mounted in a rigid housing is shown in Fig. 17.2. This is a 90 mm EFL, f/2 motion picture projection lens. All elements are air-spaced singlets because the assembly is used near a high-intensity lamp that provides a thermal loading high enough to damage the adhesive joints in cemented lenses. All metal parts are aluminum. The thick wall housing is made in two sections that thread together after the lenses are installed. Each lens is constrained by its own threaded retaining ring. No adjustments are provided so we may safely assume that mechanical tolerances are tight. Because of its rugged design, one would not expect the housings to deform significantly under gravity. Such an assembly should be designed to provide its required performance at the elevated temperature of operation.

Modular Construction

A type of construction that simplifies assembly and alignment follows modular principles. Here, alignment of the optical axis of a subassembly or assembly with respect to one or more of its mechanical reference surfaces is accomplished during its manufacture. Alignment of that axis is then automatically achieved when the module is assembled into an instrument having a corresponding interface. Some modules are designed as nonmaintainable units that are replaced rather than repaired if damaged.

Figure 17.3

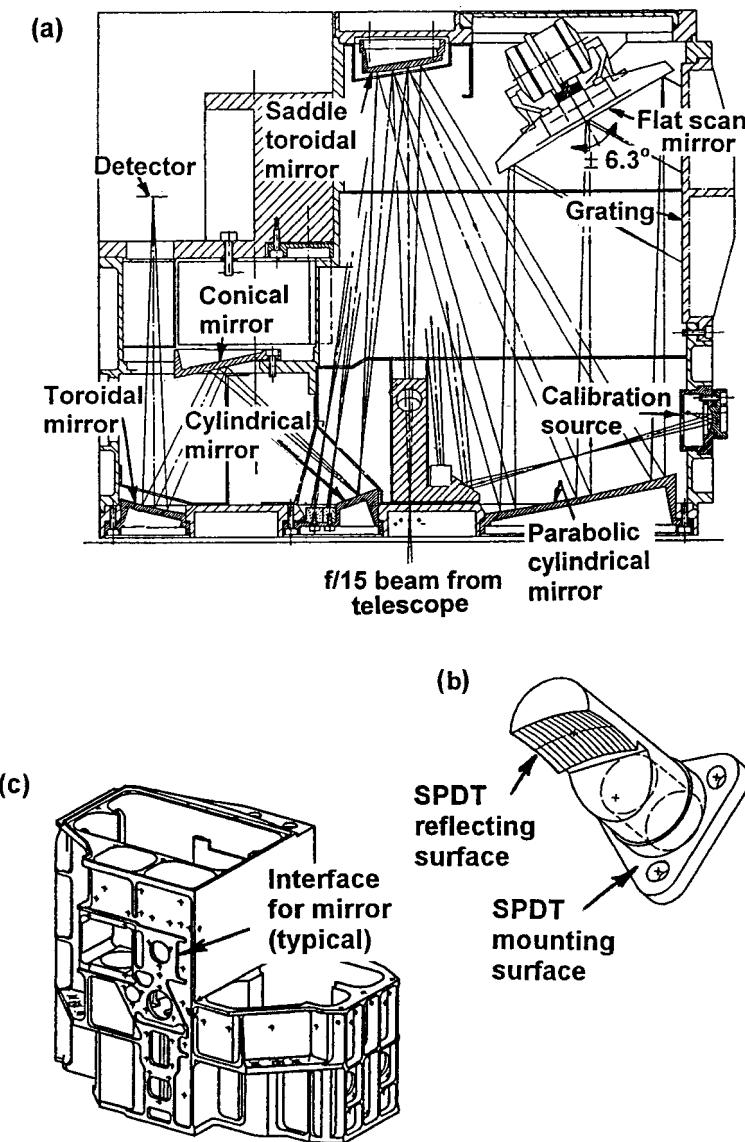
Exploded View of the U.S. Army's Binocular M19 Showing its Modular Construction.
(Adapted from Trsar et al.¹⁴)



One example is the Binocular M19 shown in exploded view in Fig. 17.3. This 7×50 binocular was designed by the U.S. Army many years ago for field use as a lighter and smaller replacement for the then standard binoculars of similar magnification and aperture.¹³ That instrument comprised two objective modules, two eyepiece modules, right and left housing modules (machined from identical thin-wall aluminum castings), a hinge pin subassembly, a neck strap and a few screws. Production quantities of each type of module were identical because all their mechanical interfaces were machined to close optical alignment tolerances after the optics were installed. A binocular could then be assembled from modules without need for alignment. Instruments could easily be repaired in the field by replacing damaged modules with new parts or with usable modules cannibalized from other unserviceable binoculars.¹⁴ For comparison, the equivalent military binocular of conventional design had as many as 300 individual parts, needed more than a dozen special tools for disassembly, and could be repaired only in a maintenance depot or factory.

The modular principle has been successfully applied in many commercial and aerospace applications. For instance, an optical instrument designed in modular fashion was the Short Wavelength Spectrometer described by Visser and Smorenburg.¹⁵ This spectrometer was part of the European Space Agency's Infrared Space Observatory. It comprised dual optical systems, each using numerous specially shaped mirrors and gratings. See Fig. 17.4a. Each of these components was created by single

Figure 17.4
Modular Construction in a Space Borne Spectrometer:
(a) Optomechanical Configuration (b) Typical SPDT Mirror Module (c) Housing Machined with Precision Interfaces for the Modules (Adapted from Visser and Smorenburg.¹⁵)



point diamond turning (SPDT) the optical and mechanical interface surfaces in precise alignment to each other on aluminum blanks. One such module is shown in Fig. 17.4b. The aluminum housing of the spectrometer is shown in Fig. 17.4c. Optical modules were installed from the outside through holes in the instrument's housing and attached with

screws to precisely machined interface surfaces on the housing. This construction ensured that the optics were located and oriented properly with only minimal adjustment needed at assembly. The savings in time and labor costs resulting from reduced alignment requirements at assembly tended to compensate for the increased cost of making the individual modules to the required precision.

Lightweight Mirror Structures

Mirrors form vital parts of many optical systems. If they are small, their weight contributions may not be particularly significant, but if large—as in astronomical telescopes—their weights may be very important system design drivers. In such cases, lightweight structures are used to form the substrates for the optical surface and its reflecting coating. A lightweight mirror may be defined as one that is lower in weight than the corresponding solid design. An efficient mirror substrate is one that achieves the required size and structural stiffness with a minimum amount of material. The design task is to put that material in the right places. Some lightweight mirror configurations are inferior with respect to stiffness than mirrors of conventional construction.

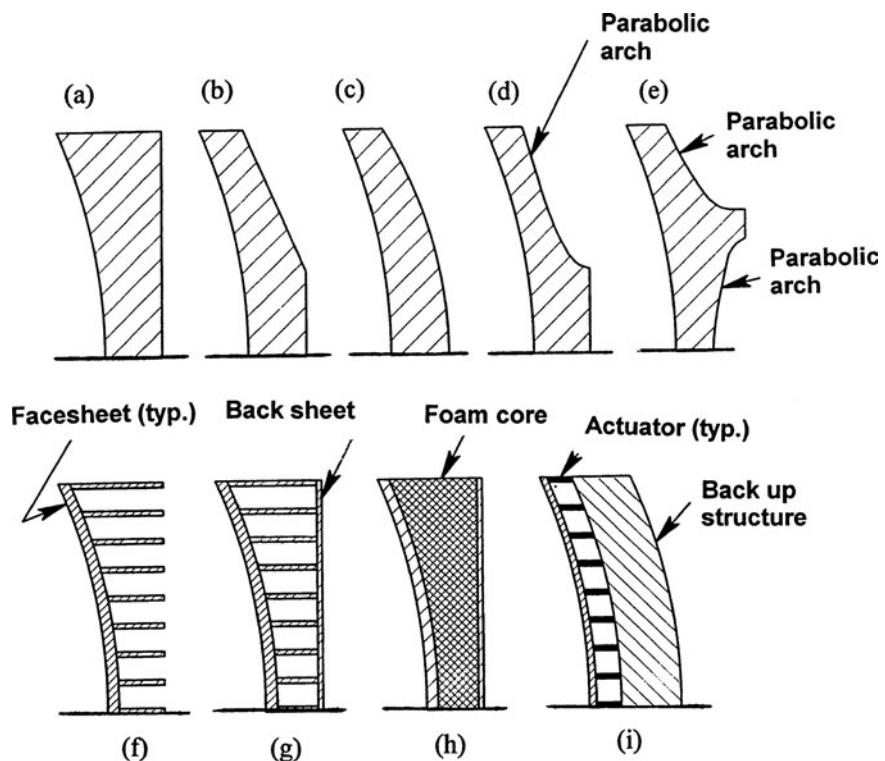
As shown by the section views of Fig. 17.5, some common techniques for reducing the weight of a concave mirror from one with a flat (nonoptical) back surface (a) are to contour the back as a cone (b), as a sphere with radius $R_2 < R_1$ (c), as a single arch (d), or as a double arch (e). Recesses can be cast or machined into the back of a solid blank or the substrate can be built from plates and strips and fused or bonded together as a sandwich with an open back (f) or a closed back (g). Foam made from glass, ceramic, silicon, or metal can serve as a lightweight core between facesheets as shown in (h).

Yet another mirror type has a thin facesheet supported by a multitude of actuators attached, in turn, to a rigid backing structure as in Fig. 17.5(i). This construction, when used with an appropriate optical surface figure sensor or image quality sensor and computer-based controls allows the mirror's surface to be driven into the optimum shape for best imagery.

The relative merits of these different mirror constructional techniques have been discussed in detail in the literature by many authors including Krim,¹⁶ Paquin,¹⁷ Vukobratovich,¹⁸ and Yoder.¹⁰

Figure 17.5

Typical Configurations of Concave Lightweight Mirrors: (a) Baseline Flat Back (b) Conical Back (c) Meniscus (d) Single Arch (e) Double Arch (f) Open-Back Sandwich (g) Closed-Back Sandwich (h) Foam Core Sandwich and (i) Adaptive thin Facesheet

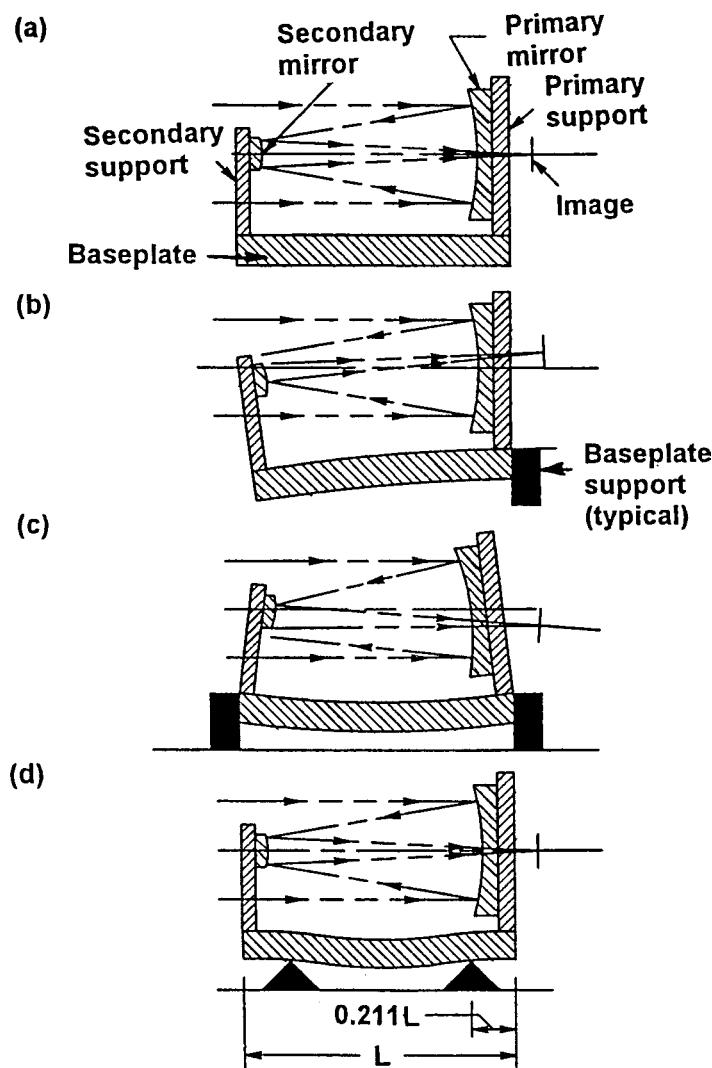


Support Structure Configurations

A simple structural element commonly used to support two or more separated optical components along a horizontal axis in a system such as a telescope or a laser cavity is a mechanical baseplate. This baseplate can have any of a variety of cross sections such as a solid rectangle, an open box, or an "I." Figure 17.6a shows schematically a side view of a simple reflecting telescope of the Cassegrain type mounted on a horizontal solid rectangular baseplate in the absence of gravity.

Actually, the baseplate must be supported somewhere and gravity may affect alignment of the optics. Figure 17.6b shows it cantilevered from one end. The opposite end sags and tilts the beam reflected from the secondary mirror. The drawing is schematic and shows angles exaggerated for clarity. If the baseplate is supported at both ends, it will sag in the middle and both mirrors will tilt as indicated in Fig. 17.6c. If the baseplate is supported at particular intermediate points (called the

Figure 17.6
 Gravity-Induced
 Deflections of a
 Cassegrain System
 Mounted on a Solid
 Baseplate Supported:
 (a) Without Gravity
 (b) Cantilevered
 (c) Supported Both
 Ends (d) Supported
 at Airy Points

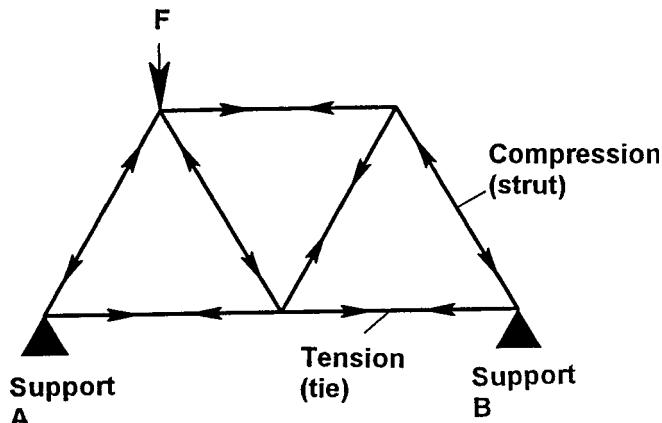


Airy points) as indicated in Fig. 17.6d, the angular optical alignment is maintained. The dimension L_A equals $\sim 0.21L$ when the weights of the optics are small compared to the weight of the beam.

Other supporting structures for optical instruments, such as tubes and boxes, also deform under gravity in ways similar to what has just been described for simple beams. To prevent excessive misalignments, these structures usually are made stiff by using materials with large

Figure 17.7

A Simple 2-Dimensional Truss Supported at Points A and B and Carrying a Load F . Tensile and Compressive Forces in the Members are Indicated



Young's moduli and/or thick walls. The latter design approach tends to increase weight.

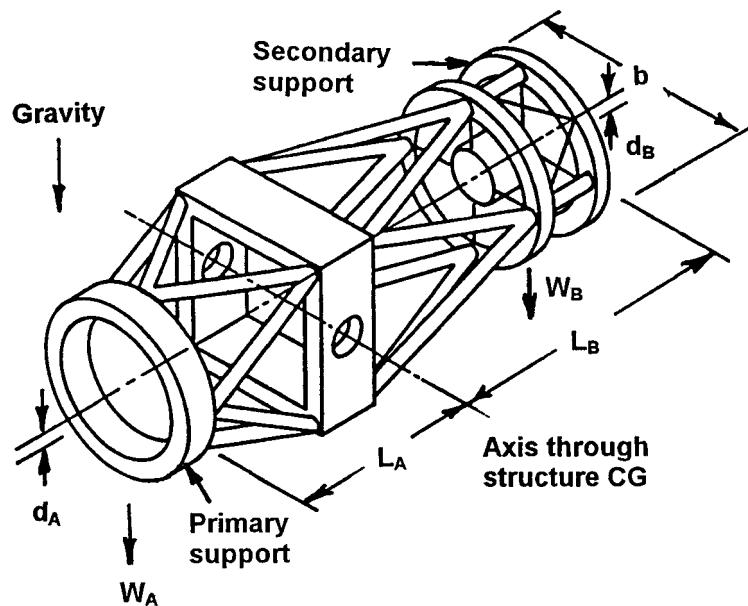
To minimize a potential weight problem without sacrificing stiffness, the structures of large optical instruments are frequently configured from triangular elements attached together to form trusses. See the two-dimensional example of Fig. 17.7. Each element of each triangle in the truss will carry axial loads well. When supported at points A and B and loaded by force F , some elements (called struts) exert compressive forces on the joints while others (called ties) exert tension. The arrows designate these conditions.

Theoretically, pins link the joints of a truss so moments and shears are not transmitted. In real life, joints usually are welded, bonded, or bolted together. Any truss-type structural design should be evaluated to determine if distortions from nonaxial forces are significant. Note that best stiffness-to-weight results from specific optimum lengths of each strut in the truss.

Proper structural design can minimize potential problems in an optical system. Consider the truss originally developed by Serrurier¹⁹ for use in the Hale 200-in. telescope on Mt. Palomar. This three-dimensional truss is sketched in Fig. 17.8. The gravitational droops d_A and d_B of the ends are the same and the axes of the primary and secondary mirrors (located at ends A and B respectively) stay parallel and aligned at all telescope elevation angles if these equations are satisfied:

$$\begin{aligned} d_A = d_B &= [W_A b/(4Ea_A)] \cdot [(4L_A^2/b^2) + 1]^{3/2} \\ &= [W_B b/(4Ea_B)] \cdot [(4L_B^2/b^2) + 1]^{3/2} \end{aligned} \quad (17.3)$$

Figure 17.8
The Serrurier Truss
Designed to Maintain
Alignment of the Primary
and Secondary
Mirrors in a Telescope
Under Variable Ori-
entation Gravity
Influence



where W is the load at A or B , E is Young's modulus and a the cross-sectional area of each truss element (assumed to be equal throughout). Dimensions b , L_A , and L_B are as shown. This design principle can be used to advantage in a variety of large optical instruments.

The Optic-to-Mount Interface

The interfaces between optical and mechanical components serve to hold the optics in their proper positions and orientations within the instrument. A lens, window, or mirror typically registers against mechanical reference surfaces and is held in place by axial and radial forces called preloads. A prism frequently is held against its references by springs. Small prisms and mirrors may be bonded in place on a mount with an adhesive.

The interfaces must constrain the optic so it is not irreversibly displaced or damaged by any specified worst-case environmental condition such as temperature, pressure, shock, or vibration. Under the more lenient operating conditions, the location and orientation of the optic must remain within allowable tilt, decentration, and despace tolerances while surface deformations and birefringence must permit full optical

system performance to be achieved. The interfaces also should provide dust and moisture seals as appropriate for the application.

Theoretically six forces would be applied to the optic to control all of its six degrees of freedom (three tilts and three translations). If those forces are applied independently and without redundancy to infinitesimal areas (points), the interface is called *kinematic*. Then, no moments can be transferred into the optic to bend it, but the stresses generated within the contact areas can be very large. A semikinematic interface is one with the same six constraints, but the contact areas are larger so the stresses are reduced. With increased areas comes the possibility of transfer of bending moments into the optic.

The mounting for a lens or window typically supports the optic by constraints applied to annular areas on the polished surfaces near the optic's rim and/or on the rim itself. Figure 17.9a illustrates the concept for one such design. Here, the radial clearance between the lens' outside diameter (OD) and the cell's inside diameter (ID) is small, typically of the order of 0.005 to 0.075 mm, so the lens cannot decenter by more than that clearance. Tilts of the lens in two orthogonal directions also are limited. Such a design may be referred to as a "rim contact" design. The arrows in the figure indicate that an axial reference feature, or constraint, is needed to register the lens axially and an axial preload P is needed to hold the lens against that constraint.

Errors made during edging of a lens to be mounted in a rim contact design can misalign the lens axis. Figures 17.10a and b show two possible errors: tilt of the rim and decentration of the rim. In both cases, the light beam is deviated from its intended path. Precision edging will control such errors.

Figure 17.9

Lens Mountings:
(a) Rim Contact
(b) Surface Contact

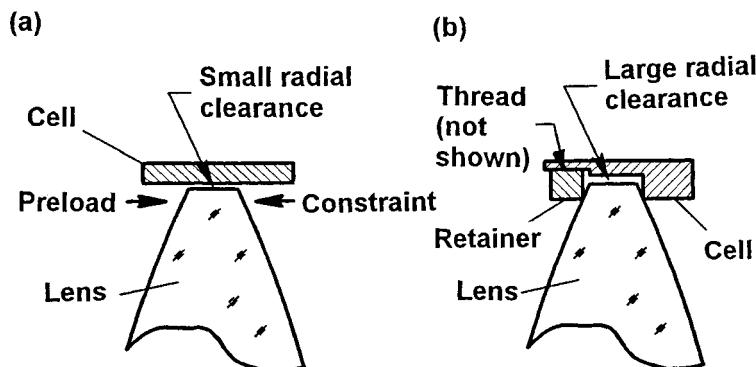
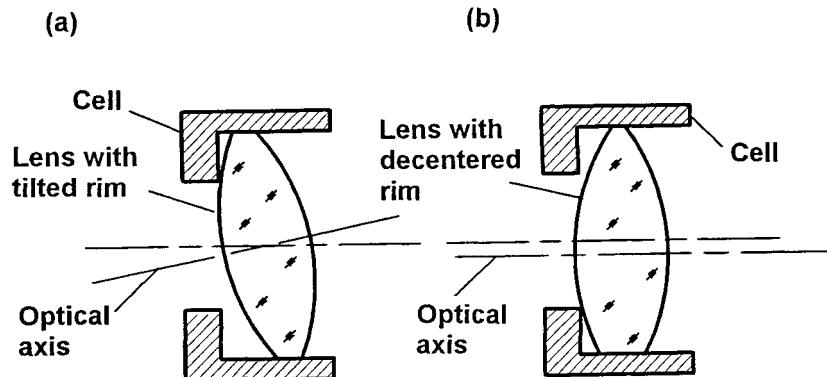


Figure 17.10

Rim Contact Lens Misaligned with Respect to the Axis of Its Mount Because of Edging Errors:
 (a) Tilted Rim
 (b) Decentered Rim



Another approach to mounting a lens is called a *surface contact* design. It has an advantage over the rim contact design in that registration of the lens occurs on its most precise surfaces—the polished ones. Figure 17.9b illustrates this type of design. The second surface of the lens registers against a shoulder integral with the cell. Axial preload is applied to the first lens surface by tightening a threaded retaining ring. The clearance around the lens rim usually is considerably larger than that shown in Fig. 17.9a. Nominally axial preload is delivered to the lens symmetrically around an annular zone on the curved lens surface. This mounting is “nonkinematic” because it can involve many contacts within this zone and within the corresponding zone at the lens-to-shoulder interface.

When the mechanical interfaces are on the lens’ polished surfaces, errors made during edging the lens are not significant because the rim is not contacted. This is indicated schematically in Figs. 17.11a and b.

Figure 17.11

Surface Contact Lens Aligned Properly in Spite of Edging Errors: (a) Tilted Rim
 (b) Decentered Rim

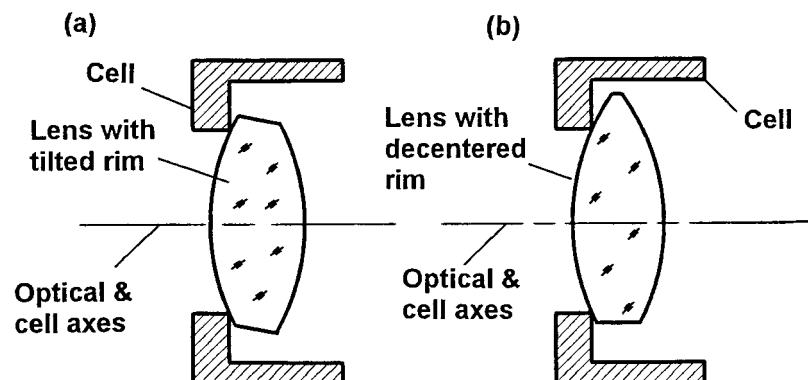
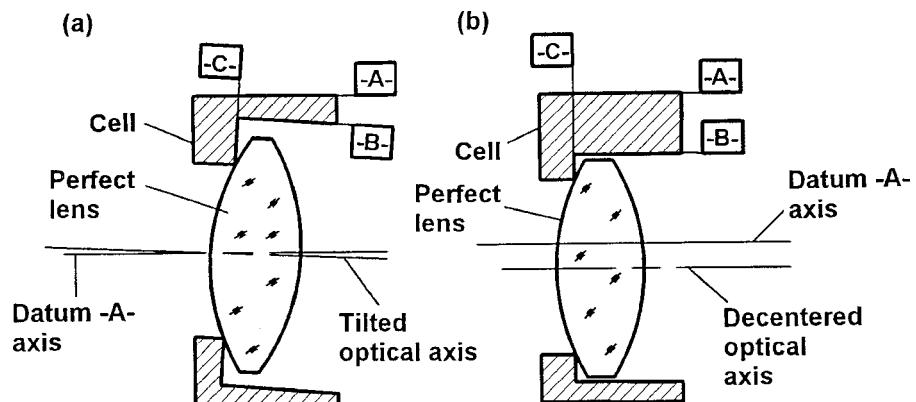


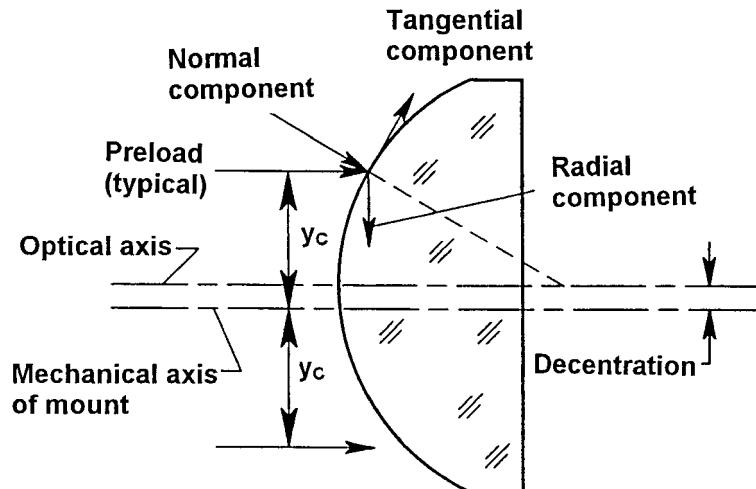
Figure 17.12
Errors in Machining Mechanical Reference Surfaces in the Mount for Rim or Surface Contact Must be Minimized to Prevent Lens Alignment Errors:
(a) Tilted Interface
(b) Decentered Interface



As shown in Fig. 17.12, errors in the mount can cause a lens to be misaligned, even if the lens is perfect. To minimize effects upon the transmitted beam, those errors must be carefully controlled. In view (a), the bore (Datum -B-) and the shoulder (Datum -C-) are tilted with respect to the cell OD (Datum -A-) while, in view (b), the bore and shoulder are decentered. Both rim and surface contact designs are affected similarly by these mount errors.

If the surface-contacted lens is initially decentered with respect to its mount during assembly, as indicated in Fig. 17.13, the preload will be more or less concentrated on one side and a radial component of that

Figure 17.13
Schematic Representation of Formation of a Radial Force Component by Axial Preload Applied to a Curved Surface of a Decentered Lens

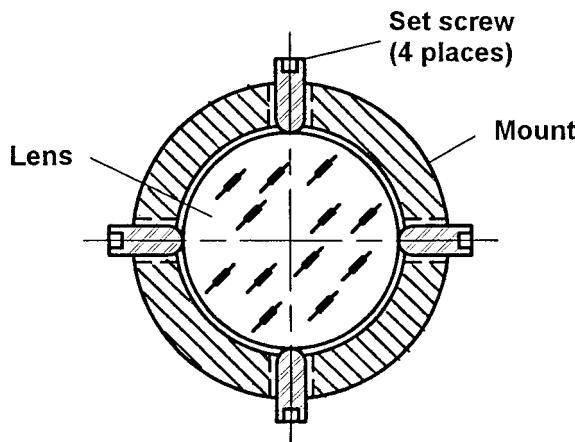


preload will tend to push the lens toward the centered condition. Friction in the interfaces on both sides of the lens at heights y_C from the mechanical axis will tend to resist lens motion. When centered, the opposing radial components balance each other. After assembly, when a centered lens is symmetrically preloaded axially, it will tend to resist decentration under vibration or shock. The ability of a given lens to self-center or to remain centered when disturbed depends largely on the radii of curvature of its surfaces and the coefficients of friction μ_1 and μ_2 at the interfaces. The longer the radii or the larger the μ_i values, the smaller is the tendency for the lens to self-center.

When, during assembly, the difference between opposing radial components of the axial preload on a given lens becomes too small to overcome sliding friction, the lens cannot center itself further. It may be possible for the centration of this lens to be improved by orienting the cell axis vertically and gently tapping the mount or vibrating the assembly to help the lens to assume its lowest possible position as the retainer is tightened. Even when this is done, a residual centration error larger than the decentration tolerance may result. To reduce this error to an acceptable value, the lens must be moved laterally by some external mechanism. The simplest such mechanism uses three or, preferably, four radially directed setscrews in "push-push" fashion. See Fig. 17.14. Each screw should have a fine thread, typically 80 threads per inch (~ 3 threads per mm), to provide adequate sensitivity.

No matter what technique for centering the lens is employed, it is essential to have a suitable method for measuring the centration error.

Figure 17.14
The Use of Two Pairs of Opposing Set Screws to Center a Lens in Its Mount



Techniques for measuring centration errors are described by several authors, notably Hopkins²⁰ and Karow.²¹ After reducing those errors as completely as possible with the chosen adjusting and monitoring equipment, the lens must be secured in the aligned position by some means that does not disturb alignment.

Independent lateral constraints are sometimes provided in a surface contact mounting for a lens or small mirror to ensure retention of proper centration of the assembled optic under vibration or temperature change. One common technique is to create shims from a suitable material with thicknesses customized for the application and to insert these at three or more equal angular intervals between the lens rim and mount wall. This essentially converts the design to a hybrid rim and surface contact form and may create an overconstrained condition at low temperatures. At high temperatures, radial contact might be lost, the shims could move, and their effectiveness would be impaired.

Vukobratovich¹⁸ described a frequently used technique, sketched in Fig 17.15, in which compliant pads are created around the lens by injecting an elastomeric sealant into radially directed holes through the mount wall at equal intervals such as 120°. A Mylar strip with three or more perforations at the locations where the pads are to be formed is inserted between the lens rim and the mount ID prior to injecting the elastomer. The sizes of the perforations will then determine the sizes of the pads. After curing, the pads locally fill the annular clearance between the lens rim and the mount and act as symmetrical compliant springs to maintain lens centration. The elastomer holds the Mylar strip

Figure 17.15

Lens Mount With Radial Constraining Pads
Created by Injecting
Elastomer through A
Perforated Mylar Strip.
(Adapted From Vuko-
bratovich¹⁸)

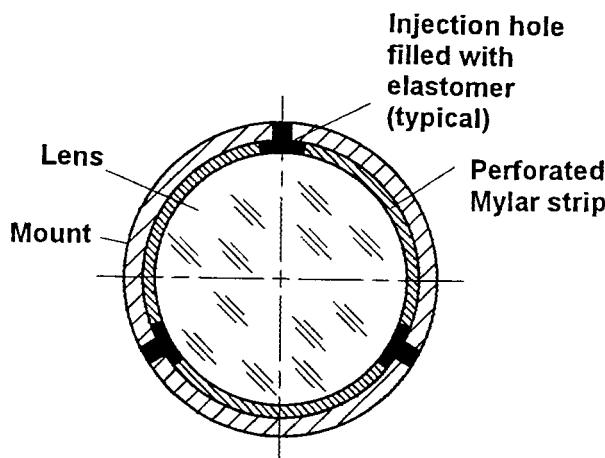
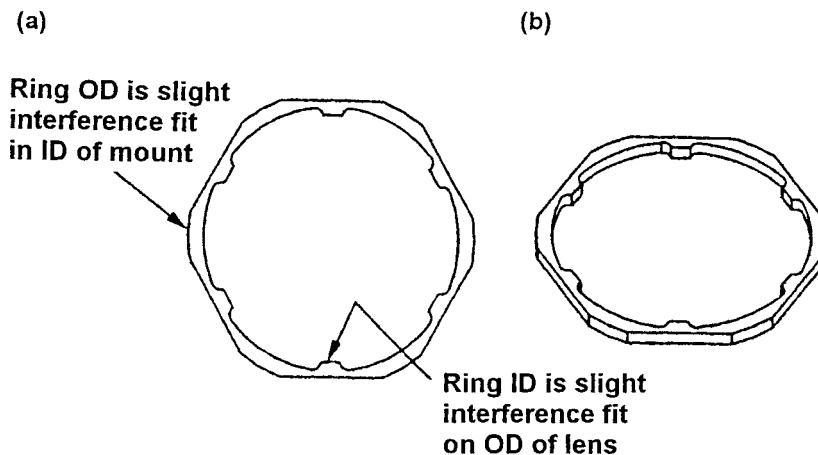


Figure 17.16

A Compliant Radial Spacer Ring Made of Vespel SP-1 Used to Center Lenses Mechanically Within the Mount ID (Courtesy of Virginia Ford, Formerly of Jet Propulsion Laboratory, Pasadena, CA)



in place. The pads should have sufficient thickness and appropriate durometer after curing to provide the required degree of resiliency throughout the specified temperature range.

Another technique, described by Ford et al.²² used a thin spacer ring made of an unfilled polyimide (Vespel SP1) to fill the annular space between the lens rim and the mount wall. See Fig. 17.16. When inserted, the ring was slightly deformed so as to load the lens rim symmetrically. Barkhouser²³ and Barrera et al.²⁴ have used more conventional mechanical springs to accomplish this function.

Establishing Axial and Lateral Preload Requirements

The nominal axial preload P (in N) needed to hold an optic in contact with its mechanical reference surfaces under vibration or shock directed parallel to the axis is:

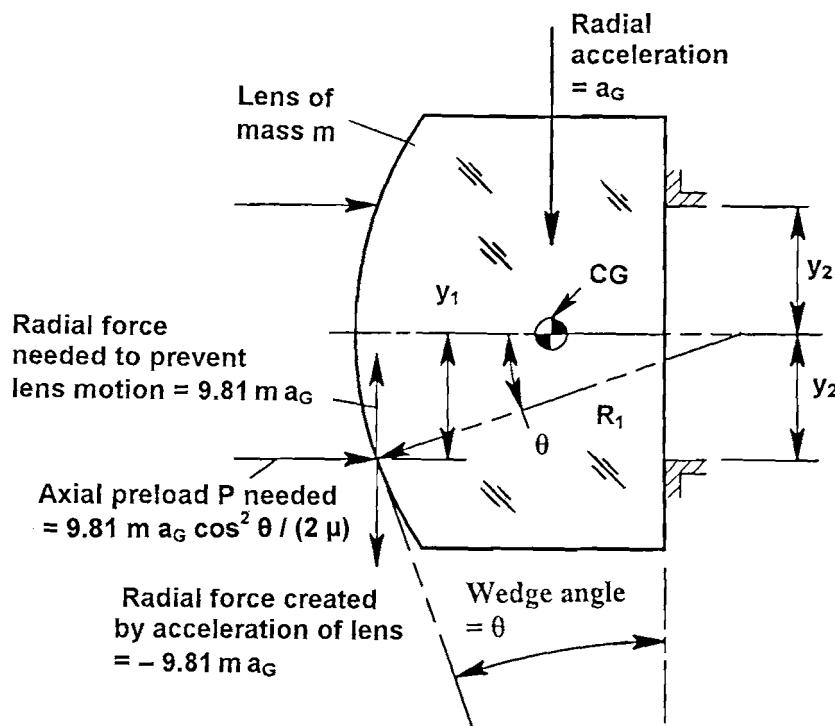
$$P = 9.81 m a_G \quad (17.4)$$

where m is the mass of the optic in kg and a_G is the acceleration factor, expressed as a multiple of nominal ambient gravity.

In a surface-contact lens mounting, the axial preload will tend to constrain the lens from decentering under lateral acceleration if that preload is applied to curved surfaces. The lens may decenter if not radially

Figure 17.17

Geometric Relationships Defining Axial Preload Required To Constrain A Lens Radially Under Radial Acceleration.



constrained. Yoder²⁵ suggesteds that this axial preload should then be estimated from Eq. 17.5 rather than from Eq. 17.4.

$$P = 9.81 m a_G \cos^2 \theta / (2 \mu) \quad (17.5)$$

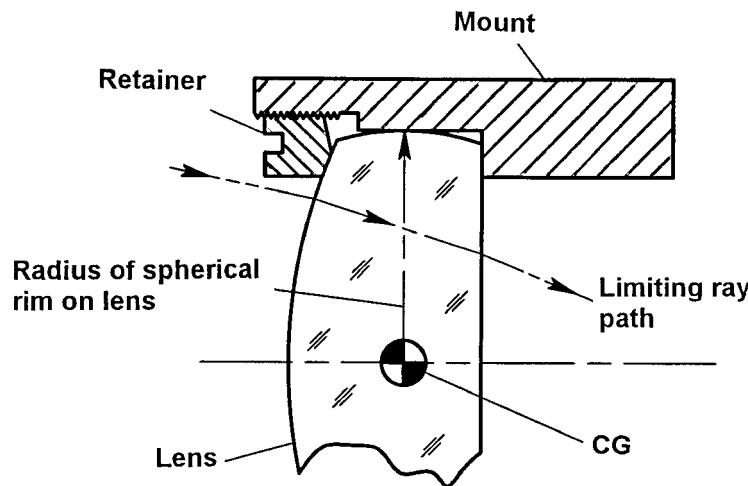
where μ is the glass-to-metal coefficient of friction and θ is the lens' geometric wedge angle at the zone of contact on both sides as indicated in Fig. 17.17. For the case of a lens with mass of 0.031 kg with θ of 36.92°, μ of 0.15, and a_G of 15, P is 4.56 N by Eq. 17.4, but 9.72 N by Eq. 17.5. If there is a possibility of loss of direct radial constraint on the lens at some temperature, the assembly preload should be increased to the greater value.

Spherical and Crowned Lens Rims

Figure 17.18 illustrates a useful design configuration in which the lens rim is fine ground as a portion of a sphere with radius equal to one-half the lens diameter (D_G). The rim then can slide into the cell at any

Figure 17.18

A Lens with a Spherical (or Crowned) Rim Facilitates Insertion into Its Mount with Minimal Radial Clearance



angular orientation. With this technique, a lens can be inserted into an opening that is only a few micrometers larger than the lens diameter without jamming in place before it is properly seated. Ideally, the high point on the rim should be in the plane normal to the axis and containing the lens' center of gravity (CG).

A variation of the spherical rim concept is a lens with a *crowned* rim. Here the rim radius is $>D_C/2$. The allowable range of tilt without jamming is smaller with this type rim than with the spherical one, but considerably larger than would obtain with a cylindrical rim. Spacers and lens cells for precision lens assemblies also are frequently made with crowned rims to facilitate assembly.

Although the spherical or crowned rim requires an extra fabrication step, the component costs are increased only slightly because the contours of rims need not to be precise. This added cost is well justified if it prevents damage to an expensive subassembly!

Interfaces for Other Optical Components

Some optical components, such as prisms and small stiff mirrors, can be supported semikinematically by a mount that constrains all six degrees of freedom (three tilts and three translations) without redundancy and

at small area contacts that serve as positional references. Forces to hold the component against these surfaces usually come from springs. Examples of such mountings are discussed later in this chapter. The mounting design needs to be carefully analyzed to make sure that the optical surfaces are not deformed by the applied forces and that undue stresses are not generated in the optic-to-mount interfaces. For conservative design, multiple mechanical surfaces touching any single surface on the optic should conform to the nominal shape of the latter surface within the tolerance allowed for distortion of the optical surface.

A nonkinematic technique frequently used for mounting prisms and small mirrors involves glass-to-metal bonds using thin layers of adhesives. These designs generally result in simple and compact packaging while providing mechanical strength adequate to withstand shock, vibration, and temperature changes.

A stress-free interface used in high precision applications to mount lenses, small mirrors, and prisms features a series of flexures between the optic and the mount. The purpose of the flexures is to isolate the optic from adverse temperature effects and to prevent introduction of mounting forces and moments that could distort optical surfaces. Optics mounted in this way may be susceptible to vibration; this aspect of the design should be analyzed.

The optomechanical interface for some lenses and many windows, shells, domes, and filters may involve “potting” the optic into its mount with an elastomeric material. This establishes the usually desired moisture, dust, and pressure seal and minimizes stresses. If sealing is not required, windows, shells, domes, and filters can be constrained with threaded retainers, flanges, or springs.

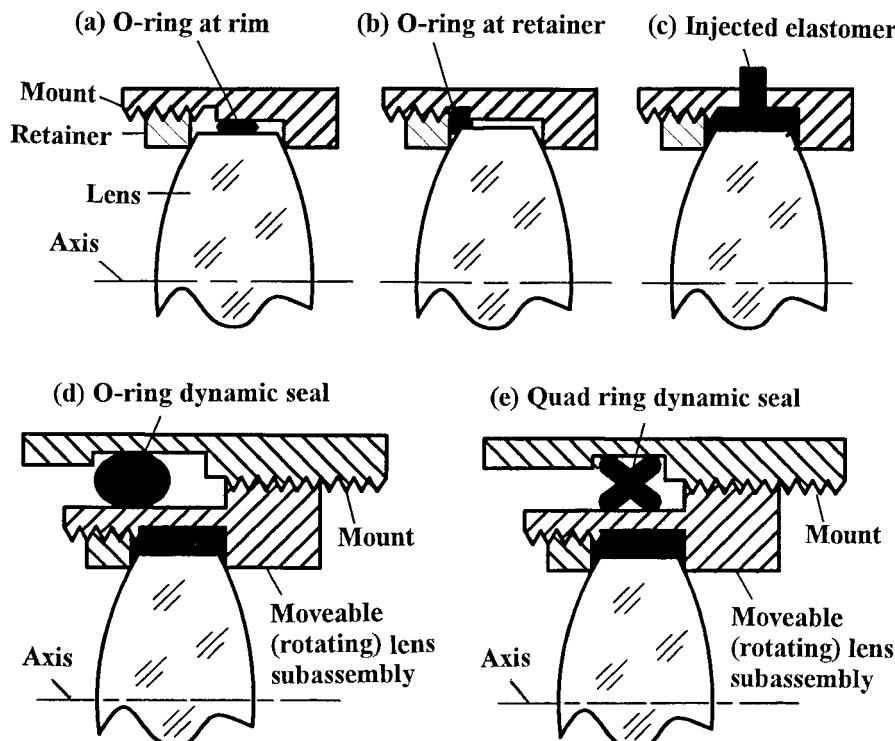
The mounts for large mirrors always are nonkinematic because those optics are thin relative to their maximum dimension and are flexible. Multiple axial and radial supports must be provided for such mirrors to minimize the self-weight deflection of the optical surfaces between supports.

Sealing Provisions

Optical instruments intended for military or aerospace applications as well as some intended for commercial or consumer uses need to be sealed against entry of moisture or other contaminants. Static sealing means include flat or convoluted gaskets, O-rings and “quad rings,” and formed-in-place elastomeric seals. Figure 17.19 illustrates three common

Figure 17.19

Techniques for Sealing a Lens Statically and Dynamically into its Mount: (a) with an O-ring at the Lens Rim (b) with an O-ring at the Retainer (c) with Injected Elastomer (d) with a Rolling O-ring and (e) with a Sliding "Quad Ring"



techniques for sealing a lens or window into a cell. In view (a), a compressed O-ring fills the radial gap between the lens rim and the cell wall. View (b) also uses an O-ring, but it is compressed against the lens face and the mount wall. Metal-to-glass contact should be retained at the interface to register the lens and limit compression of the O-ring. View (c) shows a seal injected with a syringe through several access holes in the cell wall to fill the annular region between the cell's inside surface and the lens rim. This also helps to lock the retainer.

Moving components used in some focusing mechanisms for camera lenses or eyepieces may use rolling O-rings or sliding quad-rings as dynamic seals. See views (d) and (e) of Fig. 17.19. These lens subassemblies rotate on a thread. A more complex dynamic seal using a flexible rubber bellows and translating motion is described later in this chapter.

Castings to be used as instrument housings should be impregnated with a sealant, such as a thermosetting polymer, to fill microscopic pores that otherwise can leak. This is usually done using a vacuum to remove trapped gases and cause the pores to be filled with sealant.

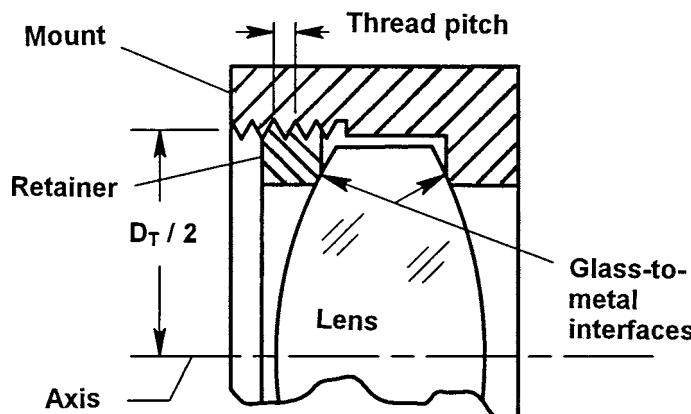
Flushing and pressurizing housings and assemblies with a dry gas (such as nitrogen or helium) are common techniques for minimizing internal damage caused by residual moisture after sealing. Some instruments are vented through desiccators and dust filters to keep their interiors clean without completely sealing them to atmospheric pressure changes. This approach is especially valuable in thin-walled devices or those with thin exposed optics because those components cannot withstand large pressure differentials.

Individual Lens Mounting Techniques

Various techniques for constraining lenses in their mounts will now be considered. The shapes of the actual interfaces between the glass and metal components are discussed in the next section.

THREADED RETAINING RING The most common technique for securing a lens into its mount is using a threaded retaining ring. See Fig. 17.20. The shoulder and the retainer contact the lens on its polished surfaces and radial clearance exists around its rim. Sets of blind holes or transverse slots are provided on the exposed face of the retainer so it can be engaged by a wrench and torqued to provide preload to the lens.

Figure 17.20
A Lens Secured in Its Mount with a Threaded Retaining Ring



Usually the lens would be centered in its mount to the required tolerance before the retaining ring is tightened. The retainer's prime function is then to clamp the aligned lens to the shoulder in this position. Manufacturing inaccuracies in location and/or angular orientation of the axis of the threaded joint between the retainer and mount or slight wedging of the retainer may, however, make the contact between the retainer and lens surface around the periphery of the lens' aperture slightly asymmetrical. If the threads have been machined so as to fit closely together, tightening the retaining ring may then move the lens out of alignment. In order to prevent this from happening, the threads should be specified to have a slightly loose fit. This is defined as Class 1 or Class 2 per *ASME Publication B1.1-2003*.²⁶ A simple *qualitative* check of the fit for a given set of parts is to insert the retainer into the mount without the optic in place, hold the subassembly near the ear, and shake it. A distinct rattle should be heard. This indicates (but does not prove) that the required fit has been provided.

The following equation can be used to estimate the axial preload in N provided by a threaded retainer when torqued in place with a wrench:

$$P = 5Q/D_T \quad (17.6)$$

Here, Q is the torque (in N-m) and D_T is the pitch diameter of the thread (in m) as defined in Fig. 17.20.

This equation is an approximation because some small factors have been ignored in its derivation and because it depends upon the values for the coefficients of friction within the thread and between the retainer and the lens surface, respectively. These parameters are never known exactly because they depend upon the initial smoothness of the surfaces in contact, how many times the retainer has been tightened and loosened, and the presence of moisture, fingerprints, and/or lubrication on the surfaces.

The same type of metal should not be used as both components in a threaded joint without lubrication or a hard coating on the contacting surfaces because the metals may then gall and seize. These surface preparations can alter the coefficient of friction in the joint significantly from that with bare metals.

The preload delivered by a threaded retaining ring loads the threads axially. A coarse thread can support more load without damage than a finer thread. Size of a thread is generally expressed in terms of its major (outside) diameter and the number of threads per unit length. The latter parameter

is the reciprocal of the thread's pitch p , which is the distance from one crest to the next. See Fig. 17.20. It has been suggested²⁵ that the minimum thread pitch on a retaining ring and its mount that is to be tightened to provide a given preload can be estimated by this equation:

$$p = 0.196 f_s P / (D_T S_Y) \quad (17.7)$$

where f_s is the desired factor of safety (typically 2 or 3), S_Y is the yield stress of the metal (in Pa), and all other terms are as previously defined. Note that the worst case preload at the extreme temperature should be used to determine thread size.

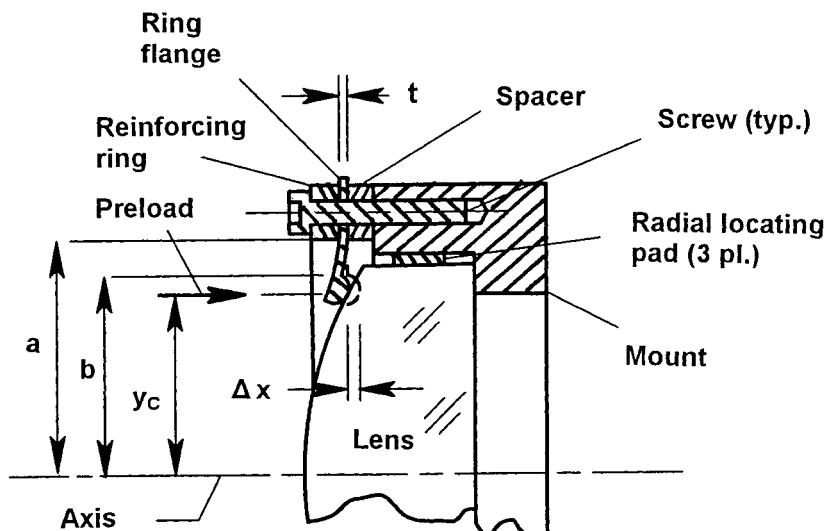
CONTINUOUS FLANGE A continuous ring-shaped flange such as is shown in Fig. 17.21 can be used to secure lenses axially. This type of constraint has been used successfully on lenses and windows >38 cm in diameter.

The flange is a thin perforated disk made of material with high yield stress. When attached to the lens mount so as to touch the lens and to be deflected by the distance Δx as indicated in the figure, it provides axial preload in accordance with this equation which was adapted from Roark²⁷:

$$\Delta x = (K_A - K_B)(P/t^3) \quad (17.8)$$

Figure 17.21

A Continuous Ring Flange Predictably Constrains a Lens in Its Mount with a Specific Preload



where:

$$K_A = \frac{3(m^2 - 1)[a^4 - b^4 - 4a^2b^2 \ln(a/b)]}{4\pi m^2 E_M a^2} \quad (17.9)$$

$$K_B = \frac{3[m^2 - 1][m + 1][2 \ln(a/b) + (b^2/a^2) - 1][b^2 + 2a^2b^2 \ln(a/b) - a^2b^2]}{[4\pi m^2 E_M][b^2(m+1) + a^2(m-1)]} \quad (17.10)$$

t is the thickness of the flexed portion of the flange between dimensions a and b as shown in the figure, and m and E_M are the reciprocal of Poisson's ratio and Young's modulus respectively for the flange material.

The material and thickness of the flange should be chosen during design to make Δx large enough to be measured accurately. Typically, one can determine this deflection to about $\pm 13 \mu\text{m}$. The deflection should then be at least ten times this value or $130 \mu\text{m}$. The thickness of the spacer under the flange can be ground at assembly to customize the deflection to the chosen value.

Bending of the flange introduces stress into that component. That stress should not exceed S_Y/f_s where S_Y is the material's yield stress and f_s is the safety factor of, say, 2. Equation 17.11 can be used to find the flange thickness that satisfies this condition.

$$t = (f_s K_C P / S_Y) \quad (17.11)$$

where:

$$K_C = [3/\pi] \left[1 - \frac{2mb^2 - 2b^2(m+1)\ln(a/b)}{a^2(m-1) + b^2(m+1)} \right] \quad (17.12)$$

Equations 17.9 through 17.12 also were adapted from Roark.²⁷

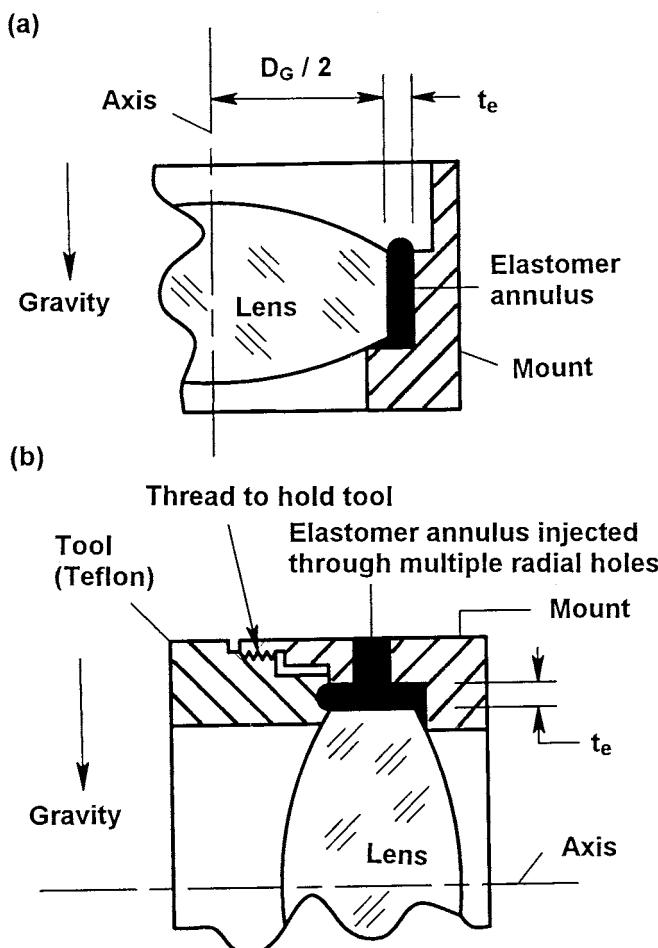
Note that a flange can be calibrated separately in terms of preload delivered as a function of deflection. When then installed in the instrument with a particular (measured) deflection, a corresponding preload can be achieved with greater accuracy than possible with a threaded retainer.

ELASTOMERIC MOUNTING A simple way to secure a lens, window, dome, filter, or small mirror in a mount uses a continuous ring of an elastomeric material, such as a room temperature vulcanizing (RTV) sealant, inserted into the gap between the OD of the optic and the ID of the mount. The lens is positioned against a shoulder and centered mechanically to the mount using shims or centered optically through

use of some auxiliary alignment measuring devices. The CTEs of the elastomer, mount, and lens respectively must follow the rule $\alpha_e > \alpha_M > \alpha_G$. The Poisson's ratio ν_e for elastomers ranges from ~0.4300 to 0.4999.

The elastomer can be inserted with a syringe into the annular gap resulting in the configuration shown in Fig. 17.22a. In this case, the lens axis should be vertical so the elastomer ring forms a symmetrical ring. The elastomer also can be injected into that gap through radially directed holes in the mount wall at several locations for a design as shown in Fig. 17.22b. In this case, the lens is best held in place with a tool that typically is made of Teflon so it can easily be removed when the elastomer has

Figure 17.22
Mounting a Centered
Lens in a Cured-in-
Place Continuous
Ring of Elastomer:
(a) Elastomer Injected
Around the Lens and
(b) Elastomer Injected
Through Multiple
Radially Directed
Holes in Mount



cured. The lens axis should be horizontal and the elastomer added through the lower holes first to allow air to escape. In both cases, the seal formed around the lens is usually adequate to prevent moisture and gas leaks.

Equation 17.13 (modified from Bayar²⁸ by Herbert²⁹) can be used to estimate t_c , the annular width of the elastomer ring.

$$t_c = \frac{(D_G/2)(\alpha_M - \alpha_G)}{[(1 + \nu_e)(\alpha_e)/(1 - \nu_e)] - \alpha_M} \quad (17.13)$$

where D_G is the lens diameter as shown in Fig. 17.22a.

The thickness so determined makes the subassembly approximately athermal in the radial direction because the elastomer tends to fill the gap between the glass and the metal at all temperatures. This occurs over the temperature range for which the material CTEs can be considered essentially constant.

Herbert²⁹ discussed two alternate equations said to give somewhat better approximations for t_c . These values are generally slightly larger than that resulting from Eq. 17.13. These variations are probably smaller than those due to the uncertainties of the parameter values entered into the equations.

An alternate design for supporting an optic uses discrete pads of elastomer symmetrically distributed around the rim of the optic rather than a complete ring. Such an arrangement will not seal the lens to its mount. The lateral dimensions of the pads will affect how the bonded subassembly responds to vibration. Optimization of these dimensions can be accomplished by finite element analysis methods.³⁰⁻³² These methods are beyond the scope of this discussion.

FLEXURE MOUNTING Extremely high performance lenses, such as those used in microlithography systems, must be aligned to extremely tight despace, tilt, and decentration tolerances. For such applications, it frequently is advantageous to attach the lens to its mount with flexures. These flexures are compliant radially, but stiff axially and tangentially.

The equal compliances of three symmetrically located flexures between the lens rim and the mount ID will keep a lens centered in its mount when the temperature changes. Because the flexures are elastic, they will allow the lens to decenter slightly during extreme (survival level) shock and vibration exposure yet return it to the correct location and orientation after these dynamic disturbances have subsided.

Figure 17.23

Mounting a Lens on Three Equally Spaced Flexures Machined Integrally into the ID of the Mount. The Lens Face is Bonded with Adhesive to Seats on the Flexures (From Bruning et al.³³)

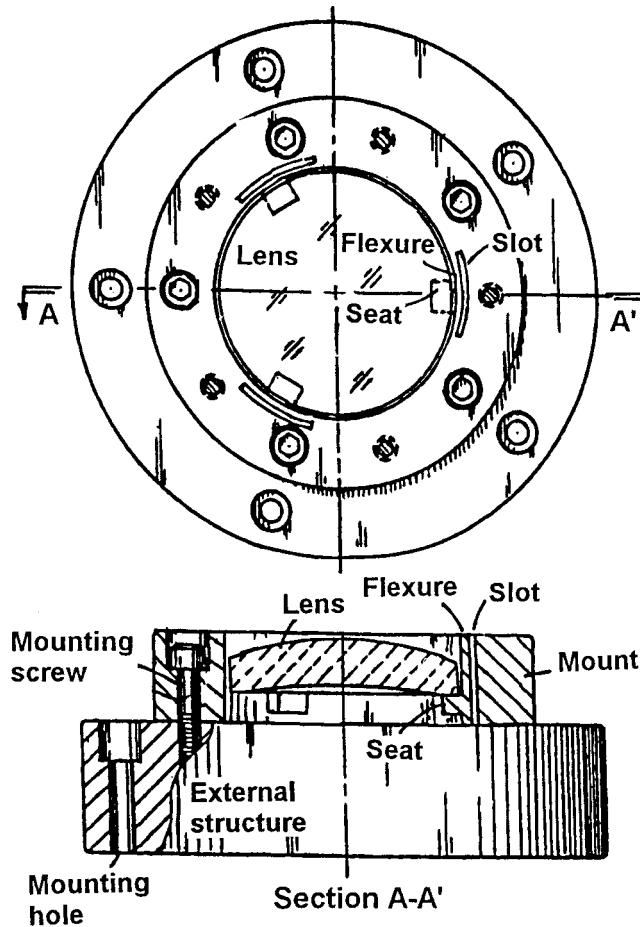


Figure 17.23 shows a design by Bruning et al.³³ in which the rim of the lens is bonded with an adhesive such as epoxy to pads on three thin flexures that are machined into the lens mount. When the dimensions of the mount and the lens change with temperature, any CTE mismatch of those parts causes the flexures to flex slightly. Since this action is symmetrical with respect to the axis and the flexures are identical, the lens stays centered.

The mount material for such a design must be chosen, in part, so the integral flexures function reliably throughout the life of the instrument. This means that it must have a high-yield stress. Stainless steel and titanium are frequently chosen for this reason.

Surface Contact Interface Shapes

In a surface mounting for lenses, each mechanical interface with lens surfaces is one of five configurations. These are described here.

SHARP CORNER INTERFACE A sharp corner interface is created in a lens mount as the intersection of a cylindrical or conical hole and a flat surface machined perpendicular to the axis of the hole. It is the interface easiest to produce and is used in a majority of optical instruments.

The sharp corner is not actually a knife-edge because such an edge is remarkably hard to produce without creating burrs. Delgado and Hallinan³⁴ described the sharp corner contact as one in which the edges of the machined surfaces on the metal part have been burnished in accordance with good shop practice to minimize defects. The resulting corners have radii on the order of 0.05 mm. This small radius surface contacts the glass at a height y_c . Figure 17.24a illustrates schematically typical interfaces on convex spherical lens surfaces while Fig. 17.24b shows typical interfaces with concave lens surfaces.

TANGENTIAL INTERFACE In this type of interface, a conical mechanical surface contacts the spherical lens surface. See, for example, Fig. 17.25. The cone half-angle, θ , is determined by the following equation:

$$\theta = 90^\circ - \arcsin(y_c/R) \quad (17.14)$$

where y_c is the height of contact and R is the optical surface radius.

Figure 17.24

Sharp Corner Interfaces between a Mount Shoulder and
(a) A Convex Lens Surface (b) A Concave Lens Surface

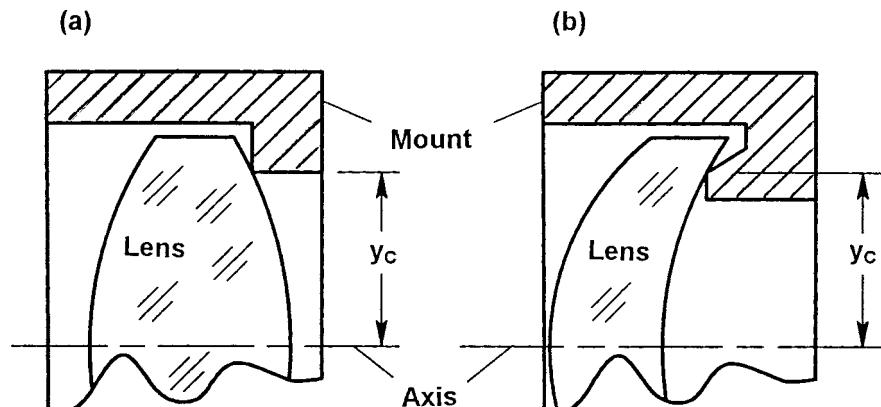
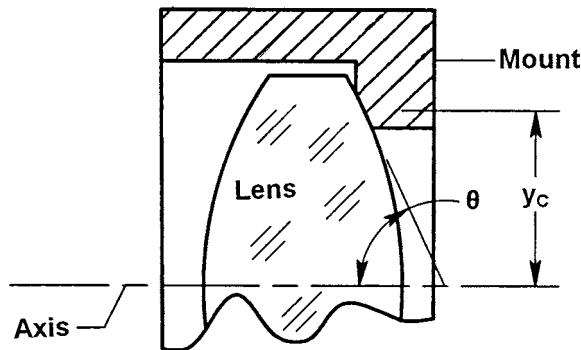


Figure 17.25
A Tangential (Conical)
Interface between a
Mount Shoulder and a
Convex Lens Surface



The tangential interface cannot be used with a concave lens surface, but it is generally regarded as the ideal interface for convex lens surfaces. It is easily made by modern machining technology. The important criterion for tolerancing the angle θ is to keep the contact near the center of the conical land. Tolerances of at least $\pm 1^\circ$ are common.

TOROIDAL INTERFACE Figure 17.26a shows schematically a toroidal or donut-shaped mechanical surface contacting a convex spherical lens surface of radius R . Figure 17.26b shows this type interface on a concave lens surface. Contact nominally occurs at the midpoint of the toroidal land in both cases. This type interface is particularly useful on concave lens surfaces. Tolerances on the surface radii of the toroids can be quite loose; perhaps varying from nominal by a factor of 2. In many cases, they can be inspected by visual comparison to templates.

Figure 17.26
Toroidal (Donut)
Interfaces between a
Mount Shoulder and
(a) A Convex Lens
Surface (b) A Concave
Lens Surface

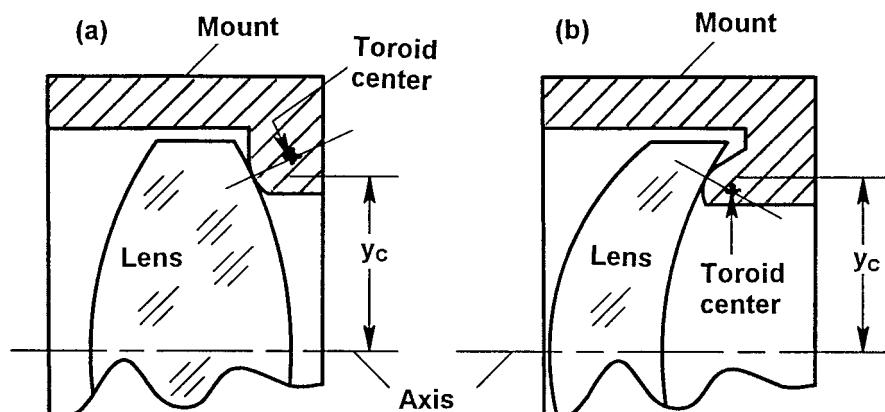
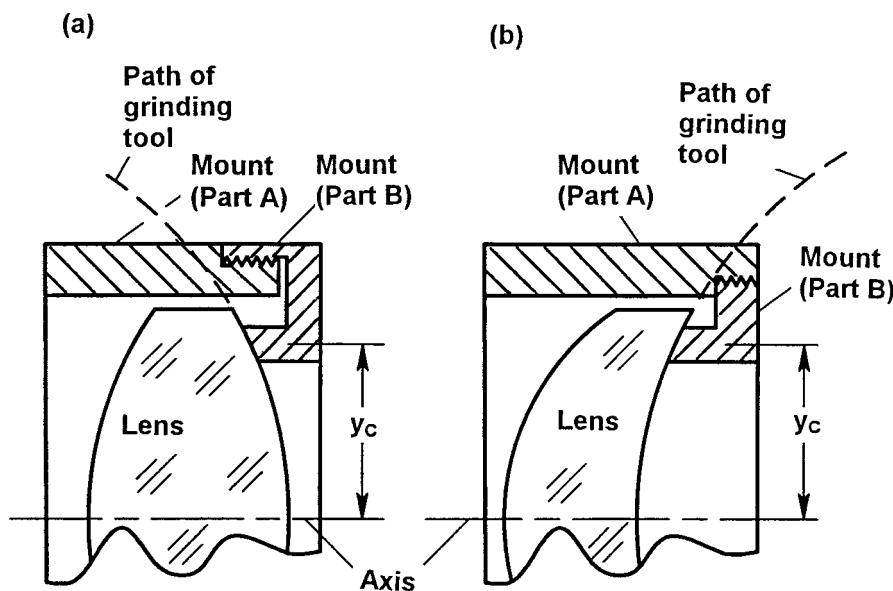


Figure 17.27

Spherical Interfaces Between a Mount Shoulder and (a) A Convex Lens Surface and (b) A Concave Lens Surface



SPHERICAL INTERFACE Figures 17.27a and b show spherical mounting surfaces interfacing with convex and concave spherical lens surfaces, respectively. This type interface has the advantage that axial forces are distributed over large areas so they cause low contact stresses under high preloads and very high acceleration loads can be survived. They also facilitate heat transfer through the interface.

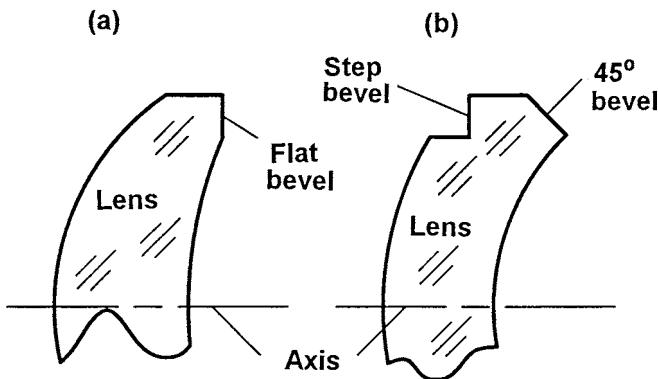
The mechanical surface that contacts the lens must be accurately ground and lapped to match the radius of the lens surface within a few wavelengths of light. The final stages of manufacture usually are done in the optical shop using tools of the same radii as those used to make the corresponding glass surface. The mount must be designed for easy access to the spherical interface surface for producing the required radius to match the lens surface. Mounts are sometimes made in two parts for this reason—as shown in the figure. Since the manufacture and testing of the mount are expensive, the spherical interface mounting technique is not often used.

INTERFACES ON BEVELS It is common optical shop practice to lightly bevel all sharp edges of optics. This minimizes the danger of chipping; such bevels are called *protective bevels*. Larger bevels (or chamfers) are used to remove unneeded material when weight is critical or packaging constraints are tight.

The concave surface of the lens shown in Fig. 17.28a has an annular bevel nominally perpendicular to the lens' optical axis. This is frequently

Figure 17.28

Lenses with Precision Bevels that Serve as Mechanical Locating Reference Surfaces: (a) Flat Bevel at a Concave Surface and (b) Step Bevel Ground into the Lens Rim at a Convex Surface. The 45° Bevel Is not Usually Used as a Mechanical Reference



called a “flat” bevel. If used as a mechanical reference, tight tolerances on perpendicularity of that bevel with respect to the lens’ optical axis must be specified for such a bevel because it determines the tilt of the lens with respect to the mount’s mechanical axis. If flat bevels are provided on both sides of a lens as, for example, on a double concave lens, those bevels must be closely parallel so that both centers of curvature of the optical surfaces can be brought to the mount’s axis within tolerance by lateral translation of the lens.

Figure 28b shows a meniscus shaped lens with a step bevel ground into the rim on the convex side. This creates a flat bevel recessed into the lens. A retainer or a spacer can be brought to bear against that surface. Perpendicularity of the flat surface to the optical axis must be tolerated adequately so the lens’ tilt meets requirements. This type bevel is useful in packaging a system with closely spaced lenses.

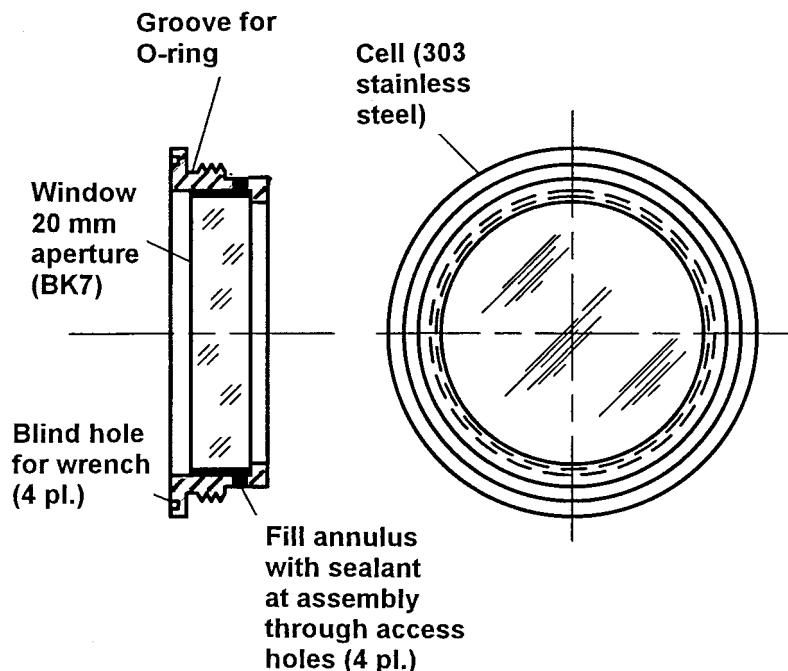
The concave surface of the lens in view (b) has a bevel angled at 45° to remove excess glass. It is not good practice to mechanically constrain the lens by contacting it on such an angled bevel because that surface may not be precisely located. Misalignment may result.

Mounting Windows, Shells, and Domes

Figure 17.29 shows a mounting design for a small window used to seal the interior of an optical system from the outside world. The application does not require high optical performance as it is used in the $f/10$ illumination path in a telescope reticle projection subsystem. Typical of this

Figure 17.29

A Window Elastomerically Sealed in Place
(Adapted from a U.S.
Army Drawing.)



type application, the surfaces need to be flat only to about ± 10 waves p-v of visible light and parallel to about 30 arcmin.

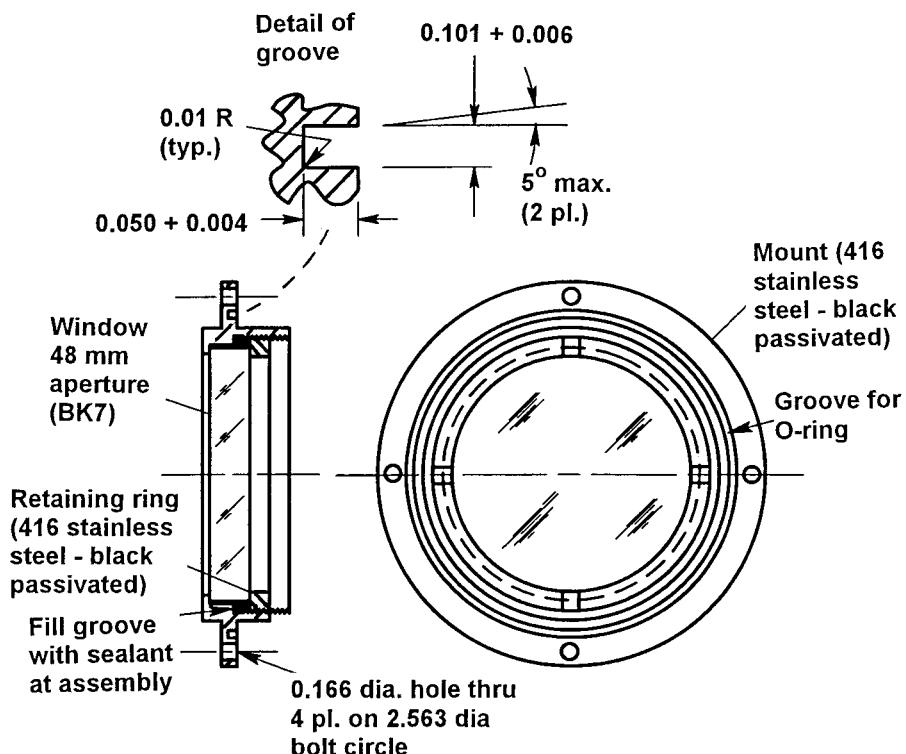
The window is bonded into a stainless steel cell with a sealing compound such as a RTV elastomer. This secures the window and forms an effective seal. Note that the window is positioned axially against a flat annular shoulder inside the cell and that the elastomer fills an annular space created by the clearance between the lens rim and the cell wall. The external thread on the cell mates with a threaded hole in the instrument housing. An O-ring seals the cell's flange to the housing. Because essentially no pressure differential exists across this particular window, it does not need to be mechanically constrained.

The window shown in Fig. 17.30 is sealed into a stainless steel cell and secured with a threaded retainer. This window is used as a protective seal in front of the more expensive objective of a high-power refracting telescope. The telescope is pressurized slightly above ambient atmospheric pressure so the pressure differential will tend to push the window against the shoulder.

Because the light beam transmitted through this window is collimated and nearly fills the clear aperture at all times, the critical

Figure 17.30

A Window Held in Place by a Threaded Retaining Ring and Sealed with Elastomer. Dimensions are Inches (Adapted from a U.S. Army drawing.)

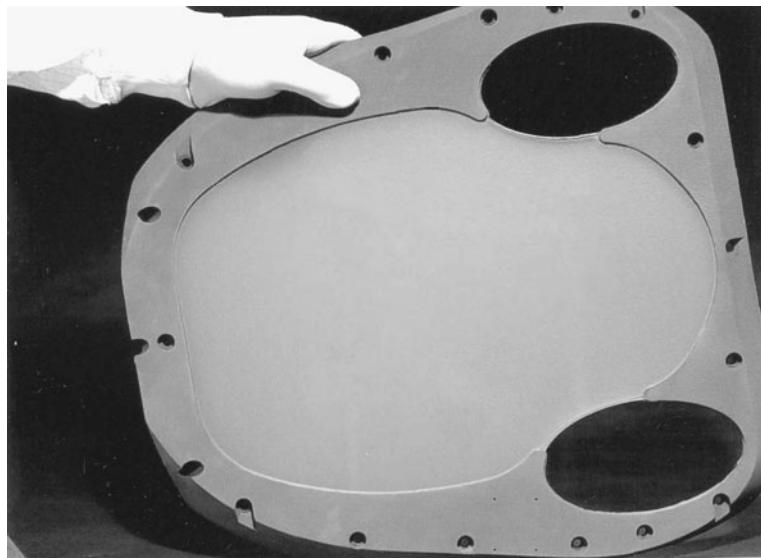


optical specifications are: transmitted wave front error ± 5 waves of spherical power, 0.05 wave p-v of irregularity for green light and 30 arcsec maximum wedge angle. Refocusing and angular alignment of the subsequent optics of the system at assembly will compensate for the window's spherical power and beam tilt. The assembly is attached to the housing of the telescope by way of a flange and sealed with an O-ring. The design of the groove for the O-ring is shown in the detail view.

Aerial reconnaissance cameras and electro-optical sensors are usually located within an environmentally controlled equipment bay in the aircraft fuselage or in an externally mounted pod. In most cases, an optical window is provided to seal the bay or pod and to provide aerodynamic continuity of the enclosure. Its quality must be high and long lasting in spite of exposure to adverse environments. Some windows for sensors used in high velocity aircraft or missiles may require cooling to counteract heating effects of the airflow across the exposed window surface. Others may need heating to prevent fogging in high humidity applications.

Figure 17.31

Photograph of a Window Assembly for an Airborne Military Application Featuring a ZnS Window (Larger Aperture) and Two BK7 Windows (Smaller Apertures). (Courtesy of Goodrich Corporation, Danbury, CT.)



The multiaperture window assembly shown in Fig. 17.31 was designed for use with a military infrared imager operating at 8 to 12 μm and a laser range finder/target designator system operating at 1.06 μm . The larger (imager) window is made of a single plate of zinc sulfide (ZnS) approximately 1.6 cm thick. Its aperture is 30 \times 43 cm. The smaller windows have elliptical apertures of 9 \times 17 cm. They are used by the laser system and are made of BK7 glass, 1.6 cm thick.

All surfaces are appropriately antireflection-coated for maximum transmission at the specified wavelengths and a $47 \pm 5^\circ$ angle of incidence. These coatings also resist erosion due to rain or ice crystal impacts at high velocity. The specifications for transmitted wave front quality are 0.1 wave p-v at 10.6 μm over any 25-cm diameter instantaneous aperture for the ZnS window and 0.2 wave p-v power, plus 0.1 wave irregularity, at 0.63 μm over the full aperture for the laser transmitter and receiver windows. The specification for geometric wedge is 66 arcsec maximum for the ZnS element and 30 arcsec maximum for the BK7 elements.

All three windows are bonded with elastomer into an aluminum mount. Retainers are not required. The bonded assembly attaches to an aircraft structure by screws through several recessed holes. The interfacing surfaces of the mount and structure must be sufficiently flat not to deform the optics or disturb their seals when clamped together.

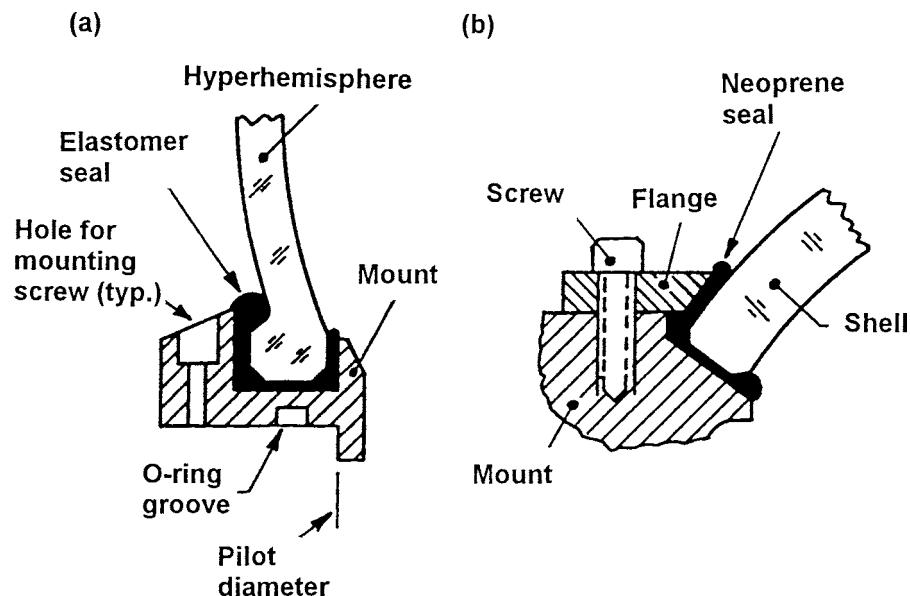
Meniscus-shaped shells are used as aberration-compensating windows in catadioptric telescope objectives such as Maksutov-type systems or in

systems requiring the ability to scan the line of sight over a large conical space. They typically are mounted in the same manner as lenses. Domes are deep shells and hyperhemispheres are domes that extend beyond 180° angular extent. Many domes are spherical, but some, intended for use on missiles, have aspheric (such as elliptical) shapes to improve their high velocity aerodynamic characteristics and minimize frictional heating. These are called *conformal domes* because they blend into the contour of the missile skin.

Domes usually are sealed with elastomer into a ring-shaped metallic mount or mechanically clamped in place through a gasket or O-ring that seals the optic. Figure 17.32 illustrates typical configurations. View (a) shows a hyperhemisphere potted with elastomer into an aluminum mount while view (b), shows a shell constrained by a flange that acts through a soft Neoprene gasket to seal the interface. More complex mountings are used on windows for high velocity missiles.¹⁰

The thickness of a window or dome is very important to survival and optical performance when the component is supporting a pressure differential. Harris³⁵ indicated that a plane parallel circular window stressed by a pressure differential ΔP_w applied uniformly over an unsupported

Figure 17.32
Schematics of Mountings for (a) Hyper-Hemisphere Potted in Place and (b) Shell Constrained with a Retainer Through a Compliant Seal [View (b) Adapted from Vukobratovich.³⁶]



aperture of diameter A_w should have a minimum thickness of t_w to provide a safety factor of f_s over the material's fracture strength S_F . This equation then applies.

$$t_w = [0.5A_w K_w f_s \Delta P_w / S_F]^{1/2} \quad (17.15)$$

where K_w is a support condition constant equaling 1.25 if the window is unclamped (as in an elastomeric mounting) and equaling 0.75 if it is clamped (as with a retainer). The customary value for f_s is 4. Typical minimum values for S_F at room temperature for some commonly used infrared window materials are given by Harris.³⁵ Elastic buckling may cause failure of a curved window (shell or dome) at stress levels considerably smaller than the material's fracture strength.

Vukobratovich³⁶ gave the following formula for the approximate OPD in mm introduced by a pressure differential into a window:

$$\text{OPD} = 0.00889 (n - 1)(\Delta P_w^2 A_w^6) / (E_G^2 t_w^5) \quad (17.16)$$

where n is the refractive index of the glass and E_G is its Young's modulus.

To illustrate these effects, consider a sapphire window with A_w of 127.0 mm and S_F of ~400 MPa if supported elastomerically and exposed to a ΔP_w of 3 atmospheres (~0.3 MPa). By Eq. 17.15, this window should be 3.89 mm thick to have a f_s of 4. We assume that this window has an E_G of 4.0E5 MPa and a refractive index of 1.684 at 3.8 μm. From Eq. 17.16, the pressurized window would exhibit an OPD of ~λ/240 at the latter wavelength.

Stress Consequences of Axial Preload

The peak tensile contact stress developed within an optic depends upon the preload, shape of the surfaces at the interface, component dimensions, and material properties. Methods for estimating this stress for typical rotationally symmetric optics (lenses, windows, and small mirrors) are summarized here. A rule of thumb tolerance of 6.9 MPa for this stress can then be applied to judge the suitability of the design.¹⁰ Stresses at mechanical interfaces with prisms are discussed later in this chapter.

The peak value for compressive stress S_C in Pa at the interface on one surface of a lens, window, or small mirror under total preload P is given by application of these equations:²⁷

$$S_C = 0.798 (K_1 p / K_2)^{1/2} \quad (17.17)$$

$$K_1 = (D_1 \pm D_2) / D_1 D_2 \quad (17.18)$$

$$K_2 = [(1 - \nu_G^2) / E_G] + [(1 - \nu_M^2) / E_M] \quad (17.19)$$

$$p = P / (2\pi y_C) \quad (17.20)$$

where D_1 is twice the lens surface radius, D_2 is twice the metal interface radius, y_C is the height of contact, ν_G , E_G , ν_M , and E_M are Poisson's ratio and Young's modulus respectively for the glass and metal. The positive sign is used with a convex lens surface and the negative sign with a concave lens surface.

To convert compressive stress S_C to tensile stress S_T we apply the following equation from Timoshenko and Goodier³⁷:

$$S_T = (1 - 2\nu_G)(S_C)/3 \quad (17.21)$$

For most optical materials, S_T is approximately $S_C/6$.

There are three interface shapes to consider: the sharp corner, the tangent, and the toroid. K_1 for a sharp corner touching a lens surface of radius > 5.08 mm is $10/\text{mm}$ and K_1 for a tangential interface is $0.5/R_{\text{LENS}}$. We apply Eq. 17.18 to obtain K_1 for a toroidal interface. Use of a convex toroid radius of $10 R_{\text{LENS}}$ as the interface for a convex lens surface and a toroid radius of $0.5 R_{\text{LENS}}$ as the interface for a concave surface has been recommended.²⁵

For example, consider a convex BK7 lens surface with R_{LENS} of 76.2 mm, y_C of 35.56 mm, P of 100 N, and a sharp corner interface on an aluminum shoulder. Then, $K_1 = 10/\text{mm}$, $p = 0.448 \text{ N/mm}$, $K_2 = 2.49 \times 10^{-5} \text{ Pa}^{-1}$ and S_C is 338 MPa. Assuming $S_T = S_C/6$, S_T is 66 MPa. This is ~ 10 times the tolerance so is unacceptable.

If we change the interface to tangential without any other design changes, K_1 becomes $0.0039/\text{mm}$ and S_T reduces to 1.30 MPa. This is acceptable. If the interface were to be changed to a toroid of sectional radius 762 mm, K_1 would become $0.0043/\text{mm}$ and S_T would be 1.37 MPa. This stress is also acceptable.

Because a tangential interface with a convex optical surface would be slightly less expensive to produce than a toroidal one while the stress

would be essentially the same; the tangential design would be preferred. For a concave optical surface, where a tangential interface cannot be used, a toroidal interface would be preferred over a sharp corner interface because it would significantly reduce stress.

Temperature Effects on Axial Preload

Optical and metallic materials for lens mountings usually have dissimilar thermal expansion coefficients (α_G and α_M) so a temperature change ΔT from the assembly temperature T_A causes a change ΔP in total axial preload P in accordance with the following relationship:²⁵

$$\Delta P = K_3 \Delta T \quad (17.22)$$

where K_3 is the rate of change of preload with temperature for the design. It is assigned a negative sign so a temperature rise decreases P . Knowledge of K_3 for a given optomechanical design would be advantageous because it would allow the estimation of actual preload at any temperature—given the assembly preload. Ignoring friction, this preload is the same at all surfaces of all optics clamped by a single retaining device. Once we predict the preload on a given lens surface at an elevated or reduced temperature and define the contact height and shape of the mechanical interface, the contact tensile stress at that interface can be estimated. This prediction is especially important for negative ΔT s because the preload can become quite large at low temperatures.

If the mount's α_M exceeds that of the glass α_G (as usually is the case), the mount expands axially more than the lens for a given temperature increase. Any axial preload P_A existing at assembly temperature T_A (typically $\sim 20^\circ\text{C}$) will then decrease. If the temperature rises sufficiently, that preload may disappear. The temperature at which the axial preload goes to zero is

$$T_C = T_A - (P_A/K_3) \quad (17.23)$$

A further temperature increase toward T_{MAX} (the maximum temperature that the instrument must survive) introduces an axial gap between the mount and lens. Imposed accelerations then may misalign

the lens. Glass-to-metal impacts under high vibration applied to the assembly while clearances exist between the lens surfaces and the adjacent mount surfaces also may damage the lens surfaces locally. To minimize this threat, it is advisable to apply sufficient preload at assembly P_A so the residual preload P_{MAX} existing at T_{MAX} will hold the lens against the mechanical interface under the maximum expected axial acceleration. The following equation defines the minimum required P_A :

$$P_A = 9.81 \text{ } ma_G - K_3(T_{MAX} - T_A) \quad (17.24)$$

where m is the lens mass and a_G is the acceleration level expressed as a multiple of g .

The factor K_3 depends upon the optomechanical design of the sub-assembly and the pertinent material characteristics. It is difficult to quantify completely and accurately, even for a simple lens/mount configuration. For instance, consider the design shown schematically in Fig. 17.33a. Here, a biconvex lens is clamped axially in a cell between a shoulder and a threaded retainer. Yoder and Hatheway³⁸ defined key mechanical changes that can occur in this design and that contribute to the magnitude of its K_3 factor. The most significant effects are:

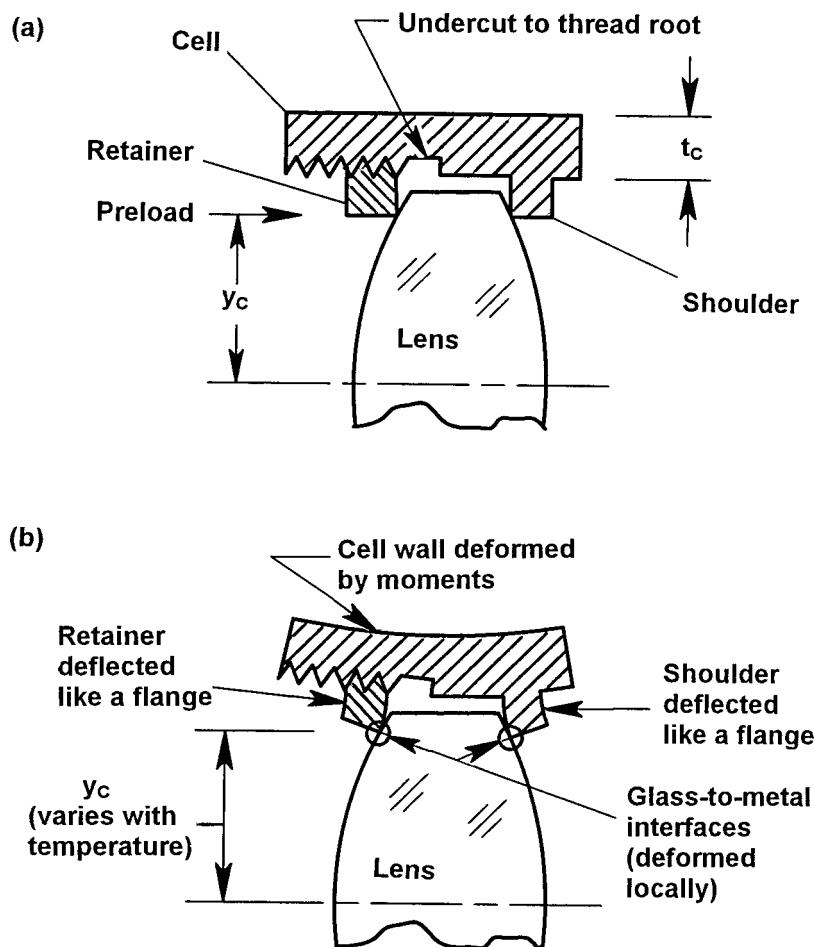
1. Bulk compression of the glass at height y_C
2. Bulk elongation of the cell wall of thickness t_C
3. Local deformations of the glass surfaces R_1 and R_2 within the opto-mechanical interfaces
4. Local deformations of the retainer and shoulder surfaces within those same interfaces
5. Flange-like deflections of the retainer and of the shoulder
6. Deformation of the cell wall by induced moments
7. Radial dimension changes of the lens and mechanical parts

Some of these changes are shown (to greatly exaggerated scale) in Fig. 17.33b. Yoder and Hatheway³⁸ discuss these effects and give applicable equations and examples of specific designs. Space limitations preclude inclusion of details here.

The latter authors also point out the common practice of many optomechanical engineers to introduce one or more axially compliant component(s) in the mounting design for a lens so as to cause K_3 for that

Figure 17.33

Exaggerated Representations of (a) A Simple Lens Mounting and (b) That Mounting Changed by Several Temperature-Related Factors Affecting Applied Preload



subassembly to become acceptably small. For example, if the retainer of Fig. 17.33a were to be made less stiff so it acts more like the continuous flange of Fig. 17.21, the adverse effect of temperature change on preload applied to that lens could be greatly reduced. Small differential axial dimensional changes of lenses and of the mount due to temperature fluctuations would, in this case, merely modify the deflection of the retainer, and hence the applied preload, very slightly. Lens mountings described by Barkhouser et al.²³ and by Barrera et al.²⁴ illustrate such techniques.

Radial Stresses and Their Variations with Temperature

Normally, some radial clearance is provided between the rim of a lens, window, or mirror and the ID of its mount, even in a rim contact type of mounting. This clearance allows the optic to be inserted into the mount. If $\alpha_M > \alpha_G$ and the temperature decreases, this clearance becomes smaller and goes to zero in some cases. Further temperature decreases then cause the optic to be compressed radially and stress to be built up in both the lens and the mount. We estimate the stress in the optic with these equations:²⁵

$$S_R = -K_4 K_5 \Delta T \quad (17.25)$$

$$K_4 = \frac{(\alpha_M - \alpha_G)}{(1/E_G) + [D_G/(2E_M t_C)]} \quad (17.26)$$

$$K_5 = 1 + \frac{2\Delta r}{D_G \Delta T (\alpha_M - \alpha_G)} \quad (17.27)$$

where D_G is the optic diameter, t_C is the mount wall thickness outside the optic, and Δr is the clearance. If K_5 is negative, the mount never touches the lens rim and no stress can develop.

The so-called “hoop stress” developed in the mount when it compresses the optic is estimated as:

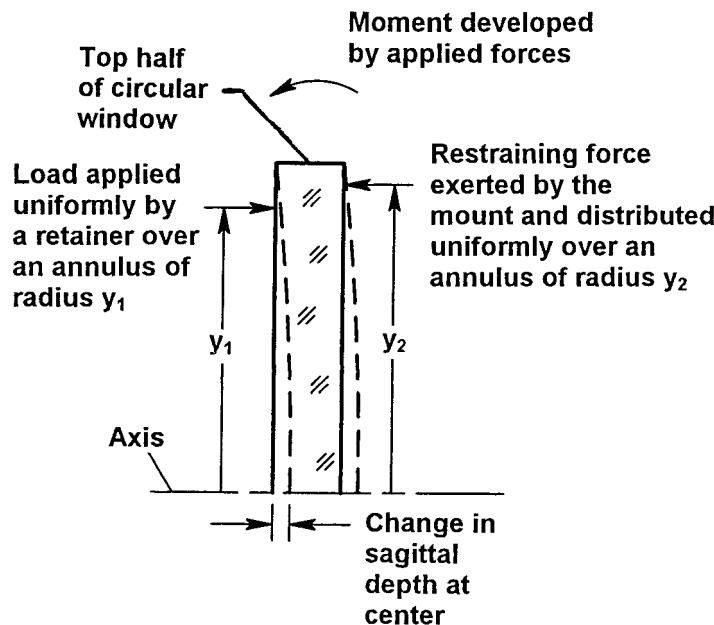
$$S_M = S_R D_G / (2t_C) \quad (17.28)$$

This stress should not exceed the yield stress of the material.

Bending Effects in Rotationally Symmetric Optics

If the axial preload and the constraint provided by the mount are not directly opposite (that is, at the same height from the axis on both sides), a bending moment is created within the optic. This moment tends to deform the optic so that one surface becomes more convex and the other surface becomes more concave, as illustrated schematically in Fig. 17.34.

Figure 17.34
Geometry Used to Estimate Stress Buildup and Sagittal Depth Change of a Window Bent by Moments from Forces Applied at Differing Heights on Opposite Faces (Adapted from Bayar²⁸.)



These deformations of the optical surfaces may adversely affect the performance of the component.

These equations allow us to estimate the change Δ_{sag} in surface sagittal depth at the center of the bent plate due to the moment:²⁷

$$\Delta_{\text{sag}} = K_8 K_9 / t_E^3 \quad (17.29)$$

$$K_8 = 3P(m^2 - 1)/(2\pi E_G m^2) \quad (17.30)$$

$$K_9 = \{(3m + 1) y_2^2 - (m - 1) y_1^2/[2(m + 1)]\} - y_1^2 [\ln(y_2/y_1) + 1] \quad (17.31)$$

where m is 1/Poisson's ratio and y_1 and y_2 are as defined in the figure.

To see if this surface deformation is acceptable, it can be compared with the tolerance corresponding to the required system performance level of the system.

Consider the following example. A 50.800 cm diameter plane parallel solid fused silica mirror with a thickness of 5.080 cm is contacted on one side by a toroidal shoulder at $y_1 = 24.130$ cm and on the opposite side by a toroidal clamping flange at $y_2 = 25.090$ cm. We assume the applied preload is 689 N. Substituting data from Table 17.1 into Eqs. 17.30 and 17.31, $K_8 = 4.380E-3$ mm² and $K_9 = 4.150E3$ mm². Then, $\Delta_{\text{sag}} = 1.387E-4$ mm = $\lambda/4.6$

for $\lambda = 633$ nm. This mirror surface deflection is probably unsatisfactory for most applications.

When an optic is bent as indicated in Fig. 17.34, the surface that becomes more convex is placed in tension. The other surface is compressed. Since glass-type materials break much more easily in tension than in compression (especially if the surface is scratched or has subsurface damage), catastrophic failure may occur if the bending effect is large. The 6.9 MPa tolerance for tensile stress given earlier applies here also.

The tensile stress S_T in a surface made more convex by bending is approximated by:²⁸

$$S_T = K_6 K_7 / t_E^2 \quad (17.32)$$

$$K_6 = 3P/(2\pi m) \quad (17.33)$$

$$K_7 = 0.5(m - 1) + (m + 1)\ln(y_2/y_1) - (m - 1)(y_1^{2/2}y_2^{2/2}) \quad (17.34)$$

where all terms are as previously defined.

For the same mirror example evaluated for deformation above, $K_6 = 722.11$ N and $K_7 = 0.45$. The tensile stress created is then 0.126 MPa. This is much smaller than the survival tolerance for glasses.

To decrease the probability of optic deflection or breakage from a bending moment in any design, the opposing contact heights should be made as equal as possible. Increasing the optic's thickness also tends to reduce this danger.

Multiple-Component Lens Assemblies

We here consider selected designs for lens and catadioptric assemblies that illustrate different types of construction. In general, these designs utilize the same principles explained earlier for mounting single optical components.

DROP-IN CONFIGURATIONS Designs in which all lenses and the interfacing surfaces of the mount are fabricated to specified dimensions within specified tolerances and assembled without further machining or adjustment are called “*drop-in*” assemblies. Low cost, ease of assembly, and simple maintenance are prime attributes of these designs. Typically,

relative apertures are $f/4.5$ or slower, and performance requirements are not particularly high. Most applications of this type design involve high-volume production and many are intended for assembly by “pick-and-place” robots. Parts are usually selected from stock at random. It is expected that a small percentage of the end items will not meet all performance requirements. Those that fail are discarded, since that is generally more cost-effective than troubleshooting and fixing the problem affecting any individual “out-of-tolerance” component.

An example is shown in Fig. 17.35a. This is a fixed-focus eyepiece for a military telescope. Both lenses (identical doublets oriented crown to crown) and a spacer fit into the ID of the cell with ~ 0.075 mm nominal radial clearance. A threaded retainer holds these components in place. Sharp-corner interfaces are used throughout to minimize cost. The accuracy of centration depends primarily upon the ability of the lenses to self-center under preload. The axial air space between the lenses depends upon the spacer dimensions, which typically are held to design values within 0.25 mm. The cell fits into a cylindrical hole in the telescope housing. It is secured, after focusing, with two setscrews and sealed with an O-ring. The outermost lens is sealed with elastomer.

A simple focusing eyepiece for a commercial telescope or binocular is shown in Fig. 17.35b. The lenses are *drop-in* assembled into a cell to form a subassembly that rotates on a thread to focus. To minimize cost for a commercial or consumer application, such an eyepiece usually is not sealed.

The axial motion Δ_E in mm required for one diopter focus change of any eyepiece is approximated as

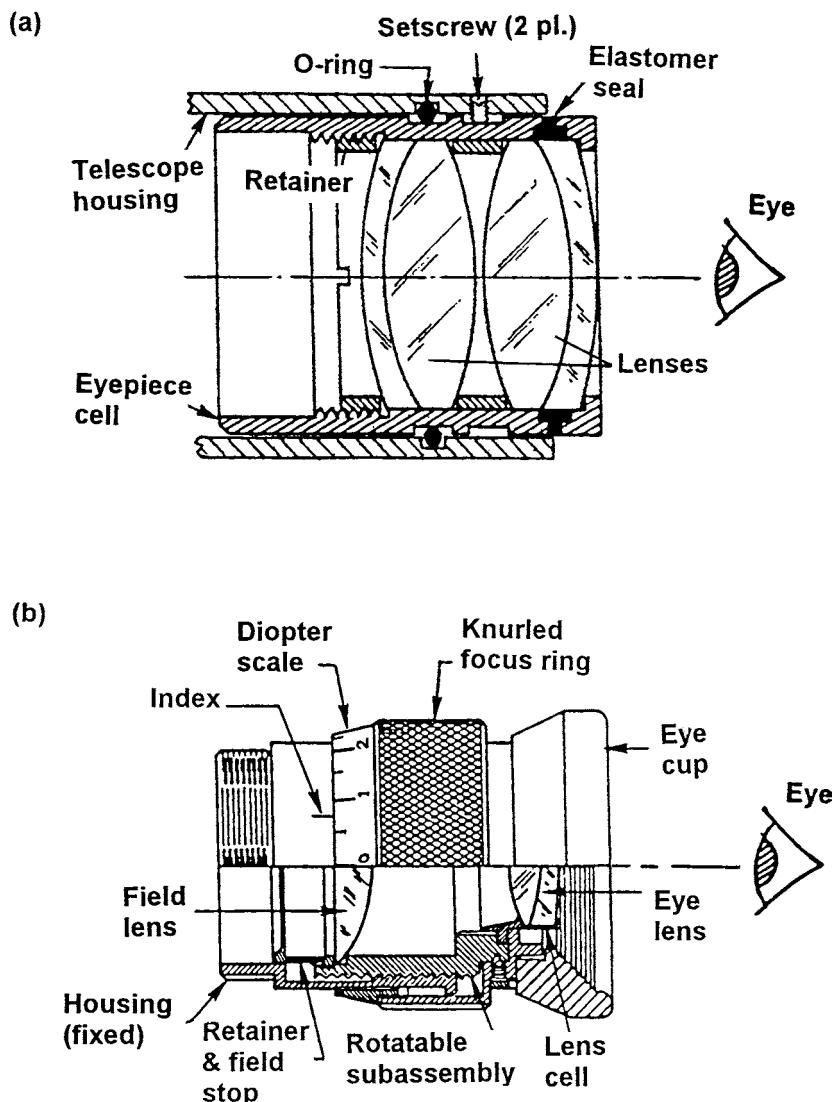
$$\Delta_E = f_E^2/1000 \quad (17.35)$$

where f_E is the eyepiece focal length in mm. Typically, adjustment of at least ± 4 diopters is required to compensate for eye accommodation errors. To avoid confusion, the rotation of the focusing ring should be limited to $< 270^\circ$.

Many eyepieces need large linear motions that are hard to provide with conventional threads. Multiple-lead threads comprising several parallel coarse threads can be used to advantage in such cases.

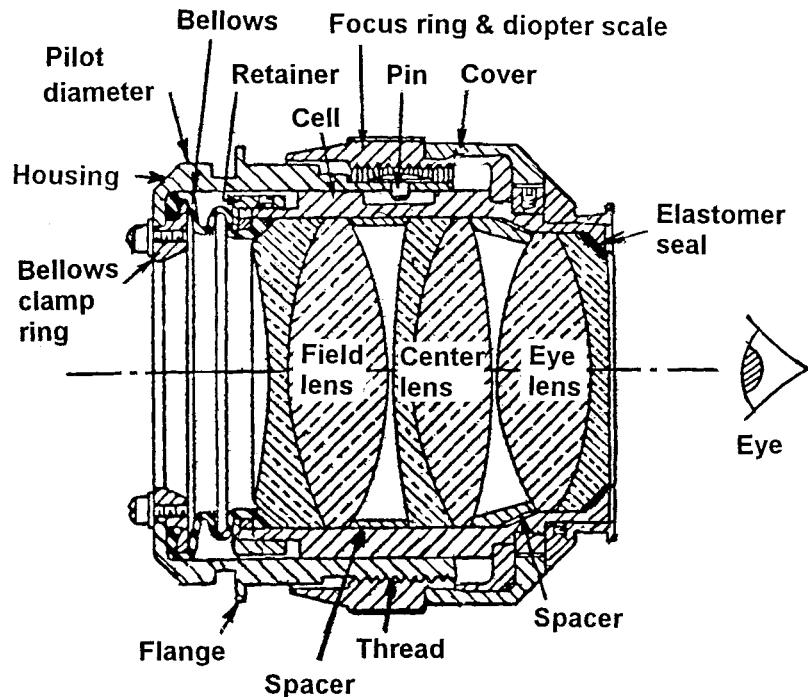
Figure 17.36 illustrates a more complex eyepiece design. Here, the lenses and spacers are assembled into a cell and constrained by a retainer. The cell slides within a housing and is driven axially to focus through action of a focus ring that is threaded onto the housing and has a flange

Figure 17.35
Eyepieces Designed for "Drop-in" Lens Assembly: (a) Fixed Focus Example Used in a Low-Power Military Telescope and (b) Low-Cost Focusing Eyepiece for Consumer Use



engaging a cylindrical groove around in the lens cell. A pin through the housing wall engages a straight slot machined into the lens cell parallel to the axis. This pin prevents the cell from rotating during focusing, so residual wedges in the lenses do not cause the line of sight to deviate. Because of this feature, the design is especially appropriate for use in binoculars where retention of collimation at all focus settings is essential.

Figure 17.36
Telescope Eyepiece
with Combined Static/
Dynamic (Rubber Bel-
lows) Seal and Lenses
that do not Rotate
During Focusing
(Adapted from
Quammen et al.³⁹)



Dynamic sealing of the focus motion is accomplished in this eyepiece with a rubber bellows between the housing and the forward end of the lens cell. This bellows also seals the field lens to the cell. The eye lens is sealed to the cell with elastomer.

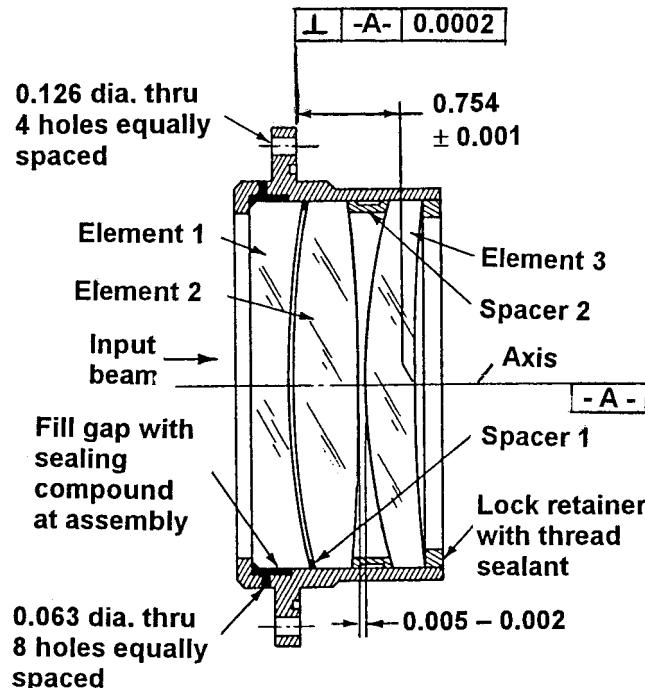
Assembly of the eyepiece to the telescope is accomplished by sliding the pilot diameter on the forward end of the eyepiece housing into a corresponding hole in the instrument until it bottoms against the flange. A ring (not shown) clamps the flange to the telescope.

LATHE ASSEMBLY CONFIGURATIONS In *rim contact* lens mountings, elements are positioned radially by close fits to the ID of the mount. Both lens and mount need to be machined precisely to circular shape and closely toleranced. The lens rim may, in some cases, be crowned for ease of assembly as discussed earlier.

A technique frequently used to mount such a lens is the so-called *lathe assembly* process. Here, the tolerance on lens OD is relaxed. The actual OD and axial thickness of each finished lens are measured and recorded. This information is kept with that lens as it goes to the assembly

Figure 17.37

Telescope Objective of Rim Contact design Made by the "Lathe Assembly" Process. Dimensions are Inches (Adapted from a U.S. Army drawing)



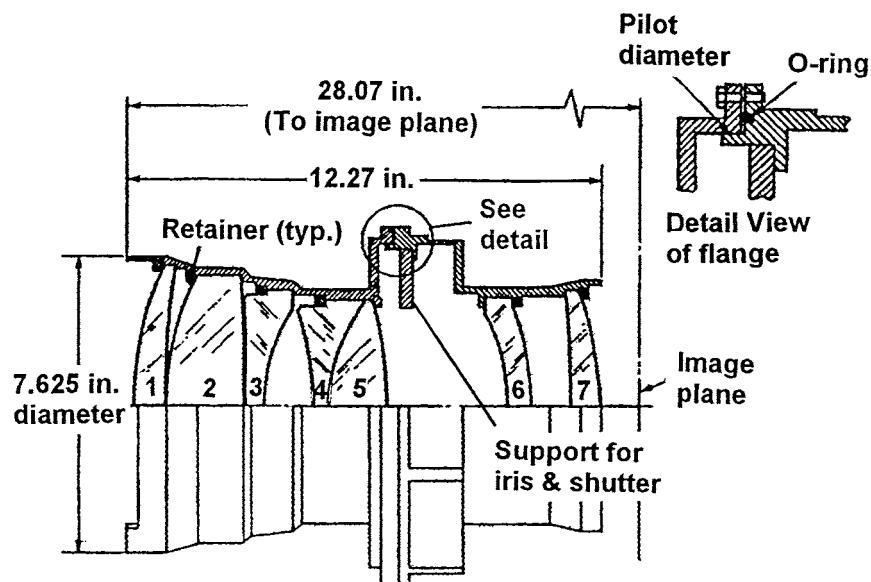
area. Mounts for that type lens will have been finish machined on all surfaces not touching the glass or affecting lens location. The remaining partially machined surfaces would then be finish machined to match the dimensions of a specific lens and to locate it properly in the assembly. These operations are usually done on a lathe. Hence, the name for the technique.

In a high-performance lens assembled in this manner, nominal radial clearance between the rim of the element and the inside of the metal part may be as small as ~ 0.005 mm. This is just sufficient for the lens to slide in place.

The telescope objective shown in Fig. 17.37 is of this type. The air spaced singlet lenses are fit into custom-made interfaces. Two spacers are used. Spacer 2 is conventional while Spacer 1 is very thin. It is machined from brass, stainless steel, or plastic shim stock and bends to conform to the glass surfaces under preload.

Figure 17.38 shows the cross section of a 61 cm focal length, $f/3.5$ aerial camera objective assembly designed for lathe assembly.²⁸ The titanium lens barrel is constructed in two parts so that a shutter and

Figure 17.38
Aerial Photographic
Lens Made by the
“Lathe Assembly”
Process Each Lens
Element is Held by Its
Own Retainer (Adapted
from Bayar.²⁸)



iris can be inserted between lenses 5 and 6 following optical assembly. Machining of the lens seats to fit the measured individual lenses begins with the components of smaller OD and proceeds toward the larger ones. Each lens is held with its own retaining ring so no spacers are required. The front and back barrels are mechanically piloted together so that their mechanical and optical centerlines coincide. The metal-to-glass interfaces are by tangent contact on convex surfaces. Concave surfaces are precision-beveled flat during centering to minimize tilt. The 3rd and 4th elements have step bevels to provide spaces for the retainers.

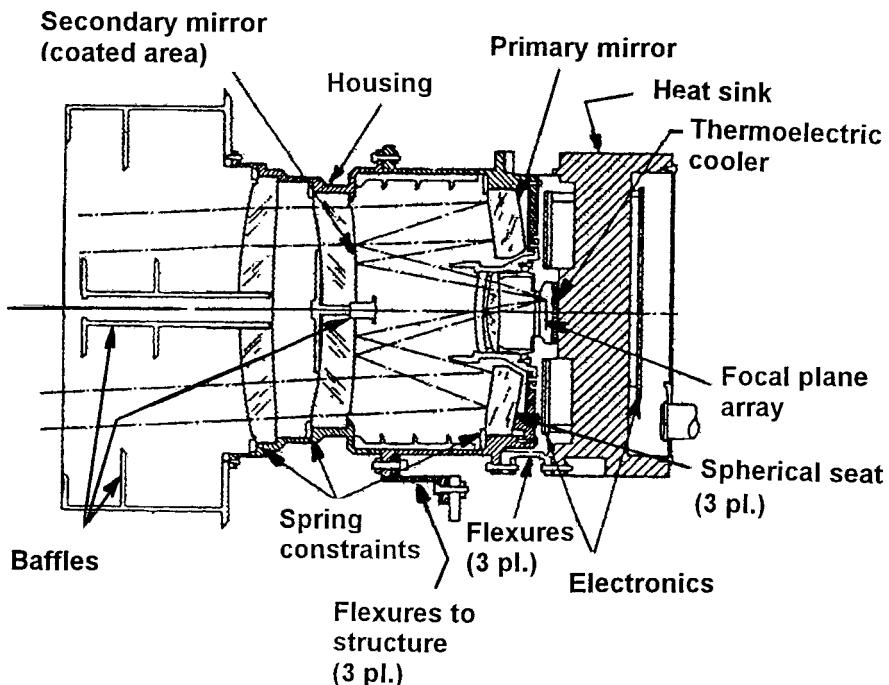
CATADIOPTRIC ASSEMBLIES

Catadioptric systems use mirrors as well as lenses as image forming optics. Two examples are described here.

A section view of a catadioptric lens developed for use as a spaceborne star sensor in a spacecraft attitude-monitoring role is shown in Fig. 17.39. This design has a focal length of 25.4 cm, a relative aperture of $f/1.5$, a field of view (diagonal) of $\pm 2.8^\circ$, and a charge transfer device focal plane assembly. The system is of the Cassegrain telescope form with the secondary mirror coated directly on the inner surface of the second large-aperture refracting element.

Figure 17.39

Schematic of a Cata-dioptic Lens Used as a Star Sensor on a Spacecraft (Adapted from Cassidy.⁴⁰)



The two larger lenses in this assembly are provided with precision annular flats that interface with shoulders in the instrument's Invar housing. They are preloaded axially with spring retainers. These elements are centered at assembly by means of radially directed setscrews (not shown) temporarily threaded through the housing wall and bearing against the rims of the lenses. After alignment, an elastomer is injected through access holes into the annular spaces between the lenses and the housing. After curing, the setscrews are removed and the vacant holes sealed.

The convex back surface of the meniscus-shaped, first-surface spherical primary mirror references against matched concave spherically ground seats in the rear cell. A flange constrains the mirror. The field lenses are lathe assembled into a cell and secured with a threaded retaining ring.

The detector subassembly, including the focal plane array, heat sink, thermoelectric cooler, and electronics, is supported from the main lens assembly by flexure blades to minimize misalignment due to differential thermal expansion of differing materials. Custom-ground spacers at each flexure attachment point fix the axial location of the array.

A classic example of a large aperture, wide field of view, catadioptric objective is the Baker-Nunn "Sattrack" camera. Developed in the mid-1950s to photograph orbiting satellites, the optical design is an enhancement of the Schmidt system. The focal length is 50.8 cm, and it operates at f/l . To prevent vignetting at the edges of the field, the spherical primary's diameter is about 79 cm.

Figure 17.40 shows, in the upper diagram, a half-section plan view of the camera; the lower diagram shows a half-section elevation view. The aperture stop of the system is very close to the center of curvature of the primary mirror, but the single correcting plate, which normally would be located there in a Schmidt telescope, is split into a triplet so as to eliminate axial chromatic aberration. The four inner surfaces of this

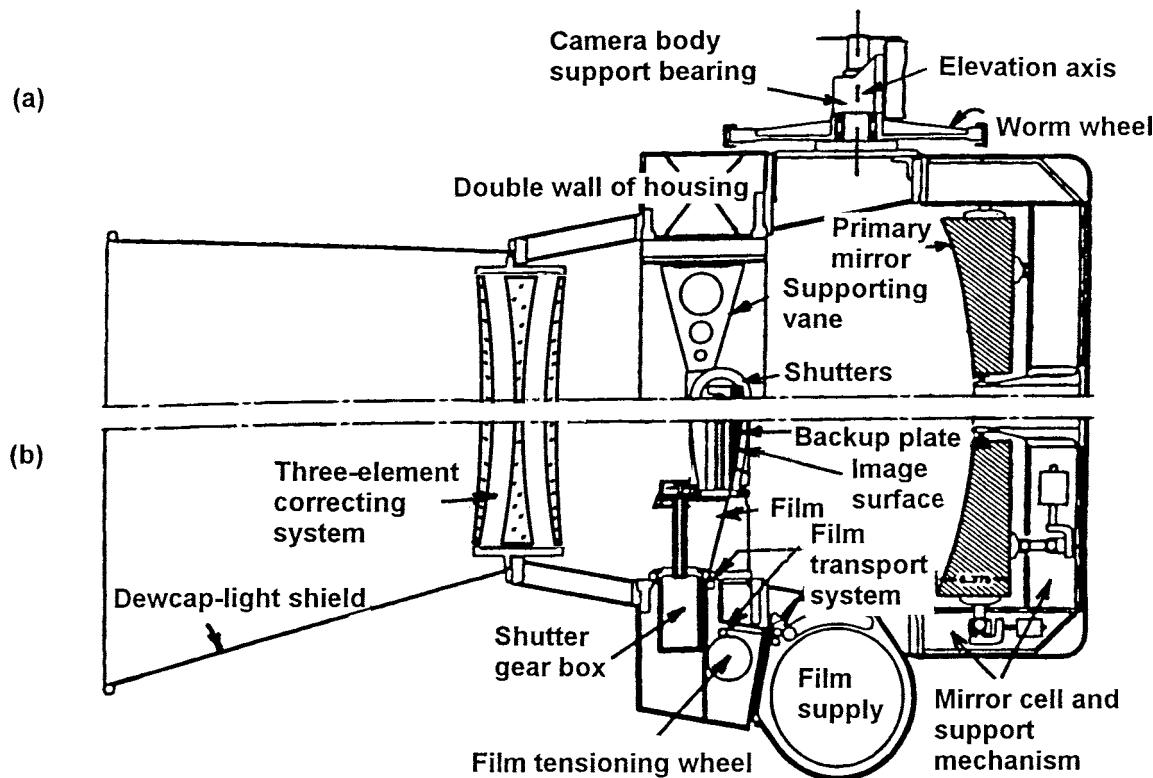


Figure 17.40

Half-Section Views of the Sattrack Camera Assembly, (a) Top View (b) Side View (Adapted from MIL-HDBK-141.⁴¹)

triplet are aspheric. The glass used in the central plate of the triplet is different from that used in the outer plates. This feature, the aspherics, and the distribution of optical power among four surfaces optimizes optical performance.

The film is transported over a cylindrically curved platen that matches the curved image. The curvature in the plane at right angles to film motion has to be zero because of the mechanical impossibility of bending the moving film into a compound curve. Consequently, the field coverage in this direction is limited to only 5°. In the direction of film travel, it is 31°. At the edges of this extreme field the focal surface departs slightly from a spherical shape so the film platen is slightly aspherical.

POKER CHIP LENS ASSEMBLIES Optomechanical assemblies with lenses mounted and aligned accurately within individual cells to form subassemblies that are then inserted as a stack into precisely machined IDs of lens barrels are sometimes referred to as "poker chips." An example is shown in Fig. 17.41. Reference surfaces are precision machined by SPDT.

An assembly featuring this type design is shown in Fig. 17.42. The lenses of this low-distortion, telecentric projection lens are aligned to the rims of their respective stainless steel cells to tolerances as small as

Figure 17.41
A "Poker Chip" Lens/Cell Subassembly. Surfaces Marked "SPDT" are Produced in one Machine Setup for Maximum Precision. The Optical Axis is Centered to the Cell OD Before Adding the Elastomer

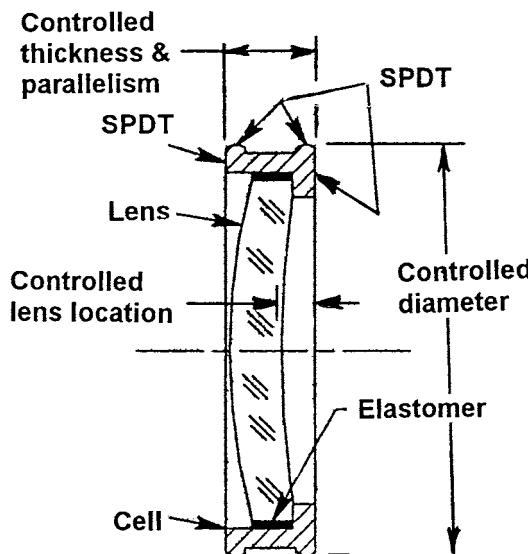
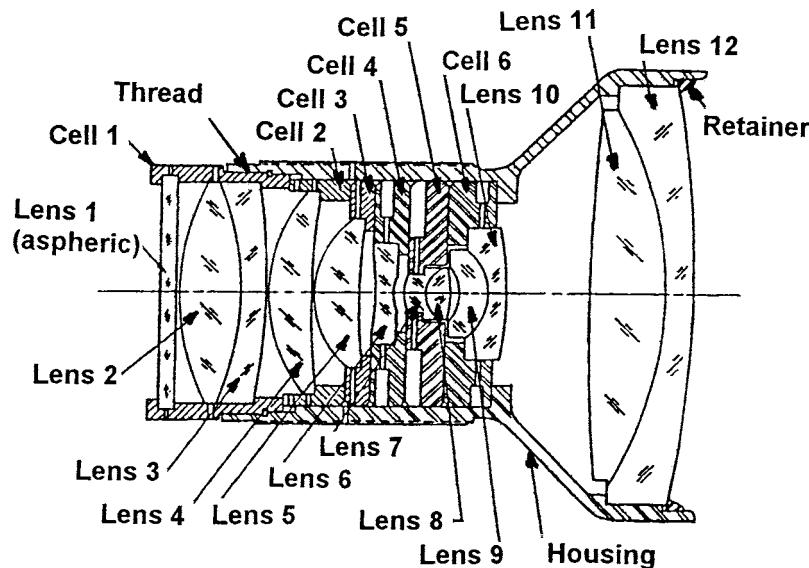


Figure 17.42

Schematic of a Low-Distortion, Telecentric Projection Lens with Poker Chip Lens Sub-assemblies (Adapted from Fischer.⁴²)



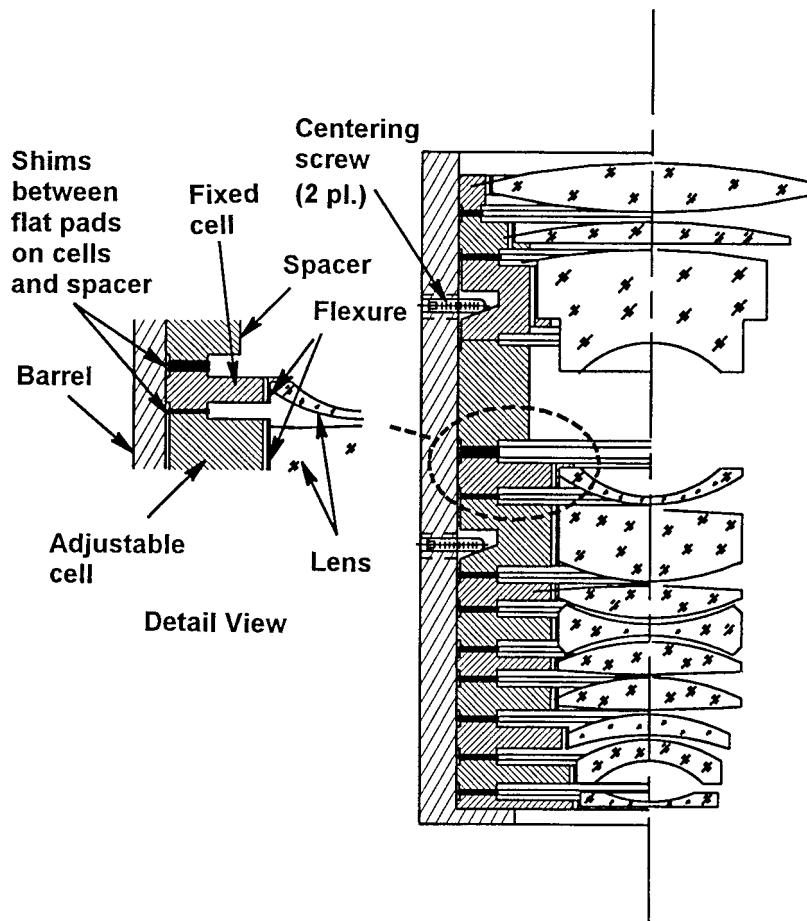
$12.7 \mu\text{m}$ of axis decentration, $2.5 \mu\text{m}$ edge thickness variation due to wedge, and $2.5 \mu\text{m}$ surface edge runout due to tilt. They are secured in place with 0.381 mm thick annular rings of epoxy injected through radial holes in the cells and cured. The cell thicknesses and front-to-back surface parallelism are machined so that the air spaces between lenses fall within design tolerances without adjustment. After the epoxy is cured, the cells are inserted into the stainless steel barrel and secured with retainers. No final adjustments are needed to achieve required performance.⁴²

Figure 17.14 illustrated a simple mechanism with four push screws for adjusting radial position of a lens as part of the alignment process. More complex versions of this mechanism are depicted in Fig. 17.43. Here, a series of 12 poker chip subassemblies are stacked inside a lens barrel. Two of the subassemblies are radially adjustable with setscrews or micrometers to optimize optical performance of the system. Typically, this operation is done in an interferometer. Spacers and customized shims between the subassemblies establish required lens axial separations.

One very important aspect of optomechanical design of a complex lens system such as this is to determine which lenses should be moved during the optimization process. One way to do this was described by

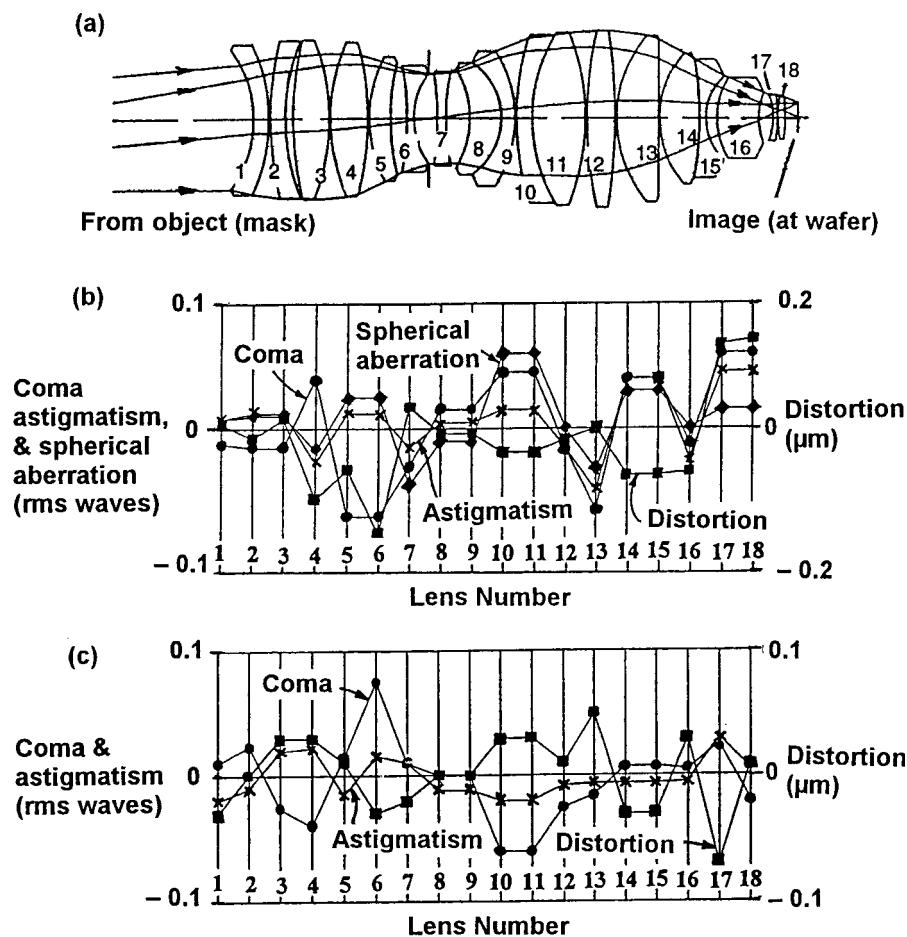
Figure 17.43

Partial Section View of a Lens Assembly with Centered Poker Chips Separated by Customized Shims to Control Air Spaces. Two Adjustable Lenses Allow Performance Optimization After Assembly (From Yoder.²⁵)



Williamson⁴³ using the system of Fig. 17.44a as an example. Here, 18 lenses relay the pattern on a mask to an image on a silicon wafer. The sensitivities of key system aberrations to 25 μm axial despace and 5 μm decentration of each lens were analyzed. These data were plotted as shown in Figs. 17.44b and c. The lens that produced the greatest effect for one aberration with minimal effect on the other aberrations was identified. For example, lenses 5 and 6 moved together would change coma significantly while astigmatism and distortion changes would be small. Similarly, lenses 8 and 9 could be chosen to modify astigmatism and lenses 14 and 15 could be used to correct distortion. All adjustments need to be clamped securely after alignment.

Figure 17.44
 (a) Optical Schematic of a $5\times$ Reduction Lens for a Microlithography Application. Views (b) and (c) Show Effects on System RMS Wavefront Aberrations and Image Distortion with Each Lens Element Shifted Axially by $25\text{ }\mu\text{m}$ and Shifted Radially by $5\text{ }\mu\text{m}$, Respectively (Adapted from Williamson.⁴³)

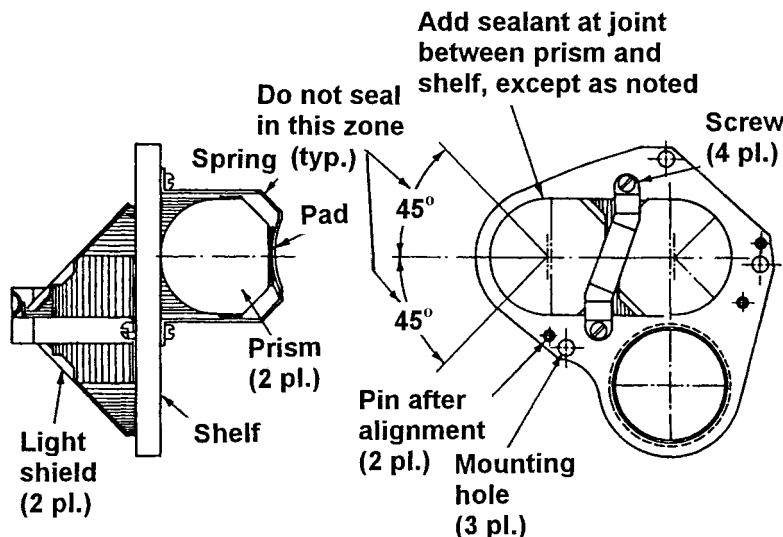


Incorporating Prisms into the Design

The suitability of any mechanical mounting for a prism or mirror depends on a variety of factors, including the rigidity of the optic; the tolerances on displacements and surface distortions; the magnitudes and locations of the preloads securing the optic; the shock and vibration forces experienced; thermal effects; the sizes, shapes, and orientations of the mounting surfaces (pads) on the mount; and the rigidity and long-term

Figure 17.45

Schematic of a Porro Prism Erecting System for a Telescope in which the Prisms are Secured by Spring Clips on Opposite Sides of a Mounting Shelf (Adapted from a U.S. Army drawing.)

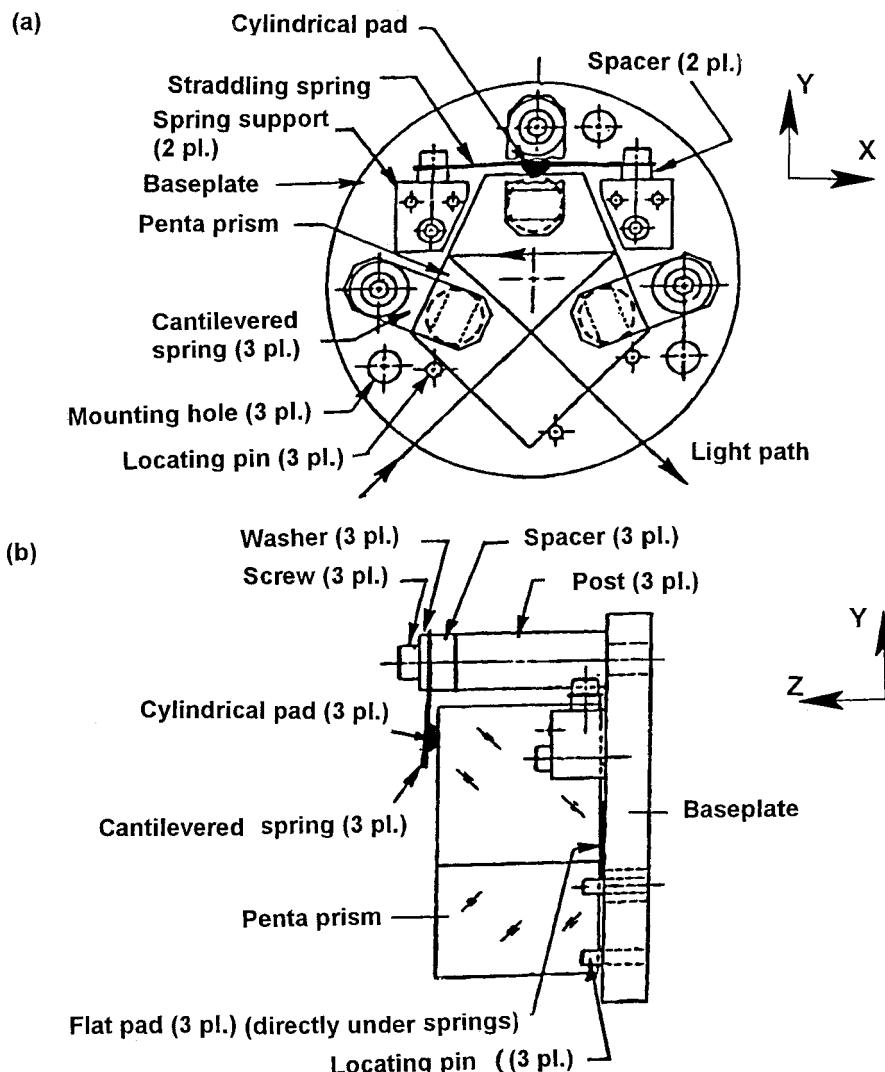


stability of that mount. Here, we describe some common designs for prism and mirror mountings.

SPRING CONSTRAINTS One technique for mounting certain types of prisms is to clamp them with springs against mounting surfaces. An example is the Porro-type image erecting system for telescopes and binoculars shown in Fig. 17.45. The prisms are optical glass with moderately high refractive index (to allow total internal reflection over the full field of view), the shelf is aluminum, the straps are phosphor bronze, the light shields are aluminum painted matte black, and the resilient pads are neoprene. The prisms fit into racetrack shaped recesses machined into both sides of the shelf. These recesses, along with sealant added along the prism edges, constrain rotation and translation of the prisms. Preload is applied to the apex of each Porro prism with a spring. This mounting is nonkinematic, but this is not a problem because the prism is very stiff.

A semikinematic mounting for a penta prism is shown in Figs. 17.46a and b. The prism is preloaded against three flat lapped pads on the baseplate by three cantilevered springs. Cylindrical pads on the springs provide favorable interfaces with the glass. In the direction parallel to the baseplate pad surfaces, the prism is preloaded by a straddling spring (with a central cylindrical pad) pushing against the prism's back face. Three short pins pressed into the baseplate locate the entrance and exit

Figure 17.46
Semikinematic
Mounting for a Penta
Prism. Preloads are
Provided by Springs.
(From Yoder.²⁵)



faces of the prism. The preloads are adjusted at assembly by customizing the spacers located under the springs to give the proper spring deflections. The following equations apply to the cantilevered and straddling springs respectively:²⁷

$$\Delta z = (1 - v_M^2)(4PL^3)/(E_M bt^3 N) \quad (17.36)$$

$$S_{B\text{cantilever}} = 6PL/(bt^2 N) \quad (17.37)$$

$$\varphi = (1 - v_M^2)(6PL^2)/(E_M bt^3 N) \quad (17.38)$$

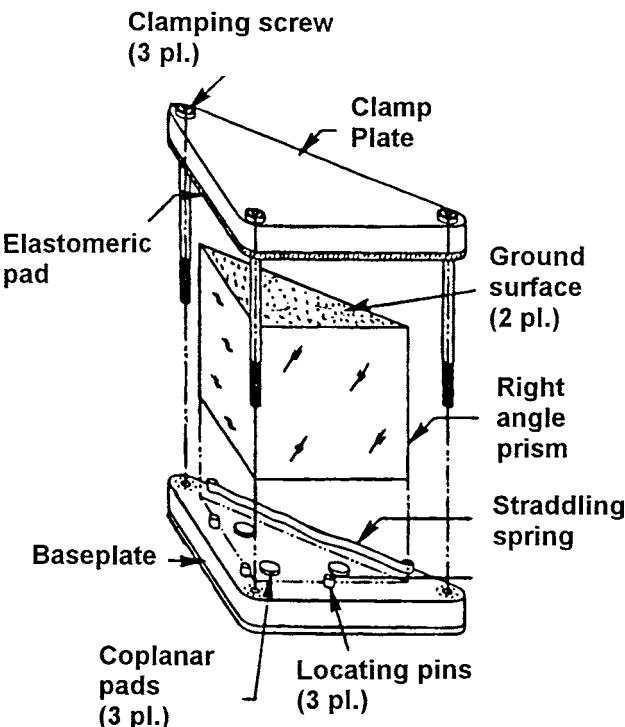
$$\Delta y = (0.0625)(1 - v_M^2)(PL^3)/(E_M bt^3 N) \quad (17.39)$$

$$S_{\text{Bstradling}} = (0.75)(PL)/(bt^2) \quad (17.40)$$

where L is the length of spring free to bend, b is the spring width, t is the spring thickness, N is the number of springs, and S_B is the bending stress in the spring. This stress should be smaller by some safety factor f_S (typically 2) than the material's yield stress.

In the prism mounting sketched in Fig. 17.47, the optic is once again clamped against three coplanar pads and three locating pins on the baseplate. Three long screws threaded into the baseplate pull the clamping plate through a resilient (elastomeric) pad to preload the prism against the three pads. A straddling spring presses the prism horizontally against the locating pins.

Figure 17.47
Semikinematic
Mounting for a Right
Angle Prism with Pre-
loads Provided by a
Compressed Elas-
tomeric Pad and a
Straddling Spring
(Adapted from Vukobratovich.¹⁸)



The resilient pad (in compression) and the three screws (in tension) provide the preload necessary to hold the prism in place under shock and vibration. It is probably safe to assume that the compliance of the pad is greater than that of the screws acting together so the latter can be ignored. To design such an assembly, the elastic characteristic of the pad material must be known. The material also should be low in outgassing and not take a permanent set when compressed for long periods of time. Yoder²⁵ described how this type of prism mounting might be designed using a material called *Sorbothane*, a viscoelastic, thermoset polyether-base polyurethane of a type that is commonly used for vibration isolation of machine tools.

BONDED MOUNTINGS Another widely used technique for mounting prisms involves glass-to-metal bonds using adhesives. This approach generally results in reduced interface complexity and compact packaging, yet provides mechanical strength adequate to withstand severe shock, vibration, and temperature changes. The technique is also frequently used in less rigorous applications because of its inherent simplicity and reliability.

The critical aspects of a glass-to-metal bond are the characteristics of the chosen adhesive, thickness of the adhesive layer, cleanliness of the surfaces to be bonded, dissimilarity of coefficients of thermal expansion for the materials bonded, lateral dimensions of the bond, environmental conditions that the bonded assembly will experience, and care with which the bonding operation is performed. Experimental verification of the choice of adhesive, bond dimensions, and bonding procedures is advisable in critical applications.

For maximum bond strength, the adhesive layer should have a specific thickness. In the case of 3M epoxy EC2216-B/A, widely used for bonding optics, experience has indicated a thickness of 0.075 to 0.125 mm to be appropriate. One method of ensuring the right layer thickness is to place spacers (wires, plastic fishing line, or flat shims) of the specified thickness at three places symmetrically located on one bonding surface before applying the adhesive. Care must be exercised to register the glass part against these spacers during assembly and curing. Another technique for creating the proper bond thickness is to mix small glass beads with closely controlled diameters into the epoxy before applying it to the surfaces to be bonded.

For design purposes, the minimum area of the bond, Q_{MIN} is:

$$Q_{\text{MIN}} = 9.81 \text{ } ma_G f_S / J \quad (17.41)$$

where m is the mass of the optic in kg, a_G is the worst case expected acceleration factor, f_s is the desired safety factor, and J is the strength of the adhesive joint in Pa (~17.2 MPa for EC2216-B/A). The safety factor should be ~4 to allow for some unplanned, nonoptimum conditions, such as inadequate cleaning during processing.

Because the dimensional changes of the adhesive bond during curing (shrinkage) and during temperature changes are related to the lateral dimensions of the bond, it is advisable not to make the bond area too large. If a large area is necessary to hold a heavy optic, the bond should be divided into a group of smaller areas such as a triangular or ring-shaped pattern of spots of convenient shape.

The following equations, reported by Vukobratovich³⁶, can be used to estimate the shear stress S_s in a bonded joint of thickness t_c and largest dimension L between two components made of materials with characteristics α_1 , α_2 , E_1 , and E_2 and thicknesses t_1 and t_2 using elastomer with given E_c and v_c when the assembly is exposed to a temperature change ΔT :

$$S_s = \frac{(\alpha_1 - \alpha_2)(\Delta T)(E_c)(\tan h(\beta L))}{2\beta t_c(1 + v_c)} \quad (17.42)$$

where

$$\beta = \{[(E_c/(2t_c(1 + v_c))] [(1/E_1 t_1) + (1/E_2 t_2)]\}^{1/2} \quad (17.43)$$

A typical bonded prism subassembly is shown in Fig. 17.48. The Porro prism is cantilevered from an aluminum bracket. Nominally, gravity is

Figure 17.48
A Porro Prism Bonded in Cantilever Fashion to a Bracket Over a Large Area for a High Acceleration Application (Adapted from Yoder²⁵)

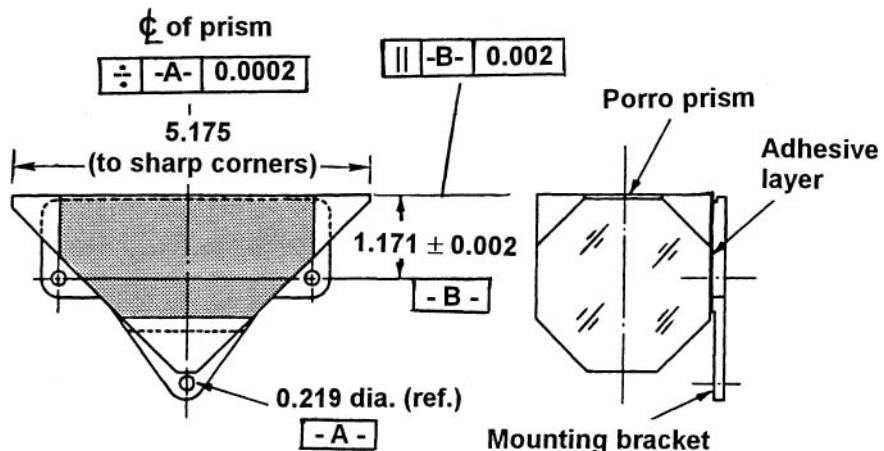
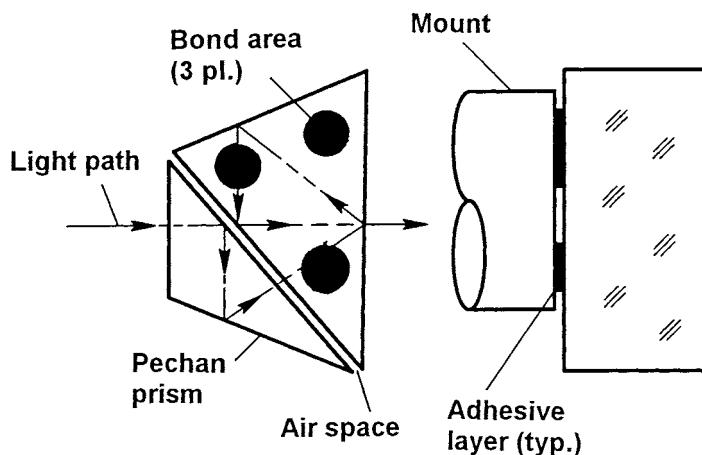


Figure 17.49

Triangular Bond Pattern on One Prism of a Two-Component (Pechan) Prism
(From Yoder.²⁵)



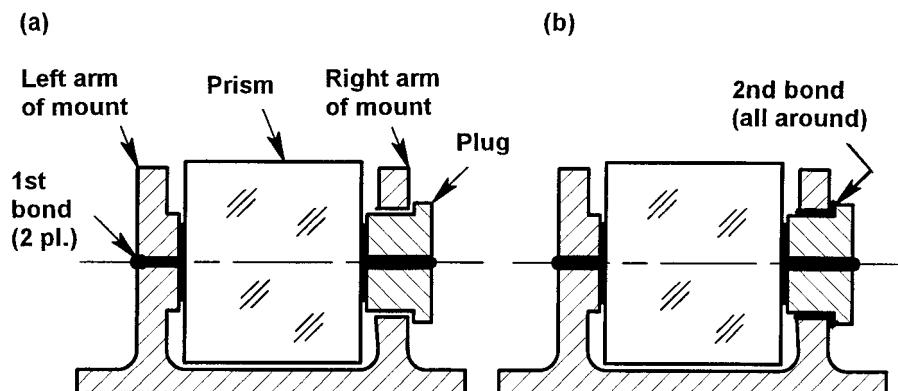
directed downward, but severe vibration is expected in all directions. The prism weighs 0.998 kg and is attached with 2216-B/A epoxy ($J = 17.2 \text{ MPa}$) over its entire side. The bond area is 3613 mm^2 . By applying Eq. 17.41, we predict that this design should survive accelerations of 1587 times gravity with a f_s of 4. The hardware was shock tested at an a_G of ~ 1200 without failure, thereby verifying adequacy of the design.

In all glass-to-metal bonds, care should be exercised during application of the epoxy to ensure that fillets of excess adhesive are not formed around the joint. Shrinkage of the epoxy along the diagonal surfaces of fillets has been known to pull chunks of glass from the optic at a low temperature!

Figure 17.49 shows a bond configuration appropriate to a two-part prism—in this case, a Pechan derotation prism. The adhesive is applied to one prism only because the ground surfaces on the adjacent components cannot be guaranteed to be coplanar. In the design shown in this figure, adhesive must not be allowed to enter the air gap because this could interfere with total internal reflection at the prism surfaces.

Bonding a prism on just one side (that is, cantilevered) may not suffice in some applications. A two-sided mounting technique suggested by Yoder²⁵ is sketched in Fig. 17.50. Here, a fixture supports the prism in proper location and orientation adjacent to a pad on one arm of the mount. A metal plug is supported near the other side of the prism within a clearance hole in a second arm of the mount. Adhesive is injected into the metal-to-glass gaps on both sides of the prism and cured. The gap around the plug is then filled with adhesive. After a second cure

Figure 17.50
Concept for Bonding a Prism so as to Support it from Both Sides (From Yoder²⁵)



cycle, the fixture is removed and the prism is then supported on both sides by bonded joints.

Mirror Mountings

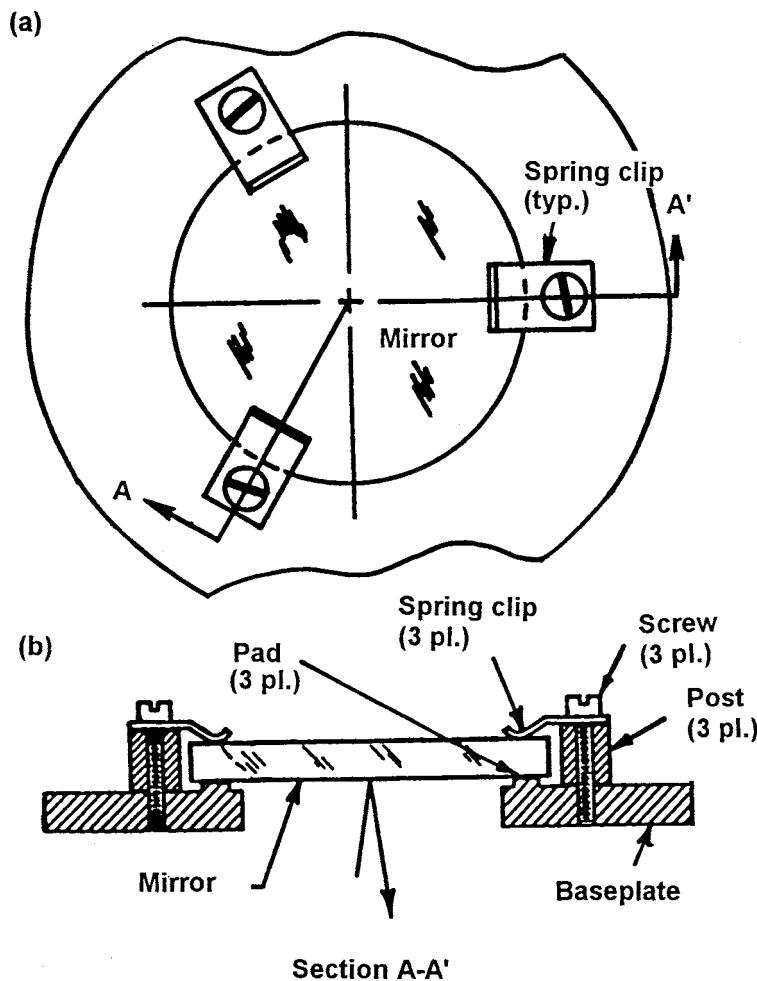
SPRING, RETAINER, AND FLANGE CONSTRAINTS Figure 17.51 shows an extremely simple means for mounting a plane-parallel mirror on a metal baseplate. The reflecting surface is pressed against three coplanar pads by three cantilevered springs. This ensures alignment of that surface. The spring contacts are directly opposite the pads so as to minimize bending moments. This design constrains one translation and two tilts. Translation and rotation in the plane of the mirror are not significant for a flat mirror. The posts are machined to the proper height for the springs to exert the needed total preload normal to the mirror. The springs would be designed in the same manner as described earlier for a prism mount. Additional springs could be used for larger mirrors needing more preload.

Circular mirrors can often be mounted in the same manner as a lens. The diameter limit for a threaded retainer type mount is set primarily by the increased difficulty of machining retaining rings to sufficient accuracy in larger diameters. Larger circular mirrors might well be held with continuous annular flanges.

BONDED MOUNTINGS Small mirrors can be mounted with glass-to-metal bonds using adhesives in much the same manner as described

Figure 17.51

Simple Spring-Clamped Mounting for a First-Surface Flat Mirror (Adapted from Durie.⁴⁴)



for bonding prisms. Figure 17.52 illustrates such a design. Equation 17.41 is used to find the proper minimum bond area. First-surface mirrors with dimensions up to about 15 cm can be bonded directly to a mechanical support. The ratio of largest face dimension to thickness of the mirror should be $< 6:1$ in order for adhesive shrinkage during curing and/or at extreme temperatures not to distort the optical surface.

Previously, we described the use of annular rings of elastomeric material to secure the rims of lenses into cylindrical mounts. This technique can just as well be applied to small mirrors. Figure 17.53 illustrates such a mounting for a small convex mirror.

Figure 17.52
Typical Bonded Mounting for a Small First-Surface Flat Mirror. A Diameter-to-Thickness Ratio of $<6:1$ is Needed to Resist Mirror Deformation. (Adapted from Yoder.²⁵)

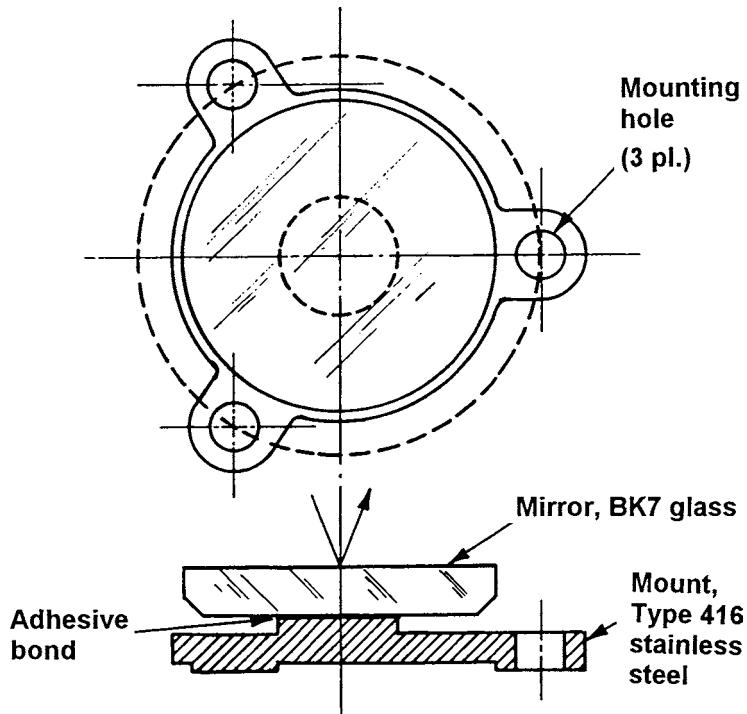


Figure 17.54 shows an elastomeric mounting for a fused silica mirror of 4.5 in. (11.4 cm) diameter supported in a Kovar cell with 12 Dow Corning 6-1104 silicone pads. Mammini et al.³² demonstrated that the fundamental frequencies of the piston and tip/tilt modes of this mirror subassembly varied with pad diameter and that the pads had to have at

Figure 17.53
Schematic of a Convex First-Surface Mirror Constrained in a Cell with an Elastomeric Ring

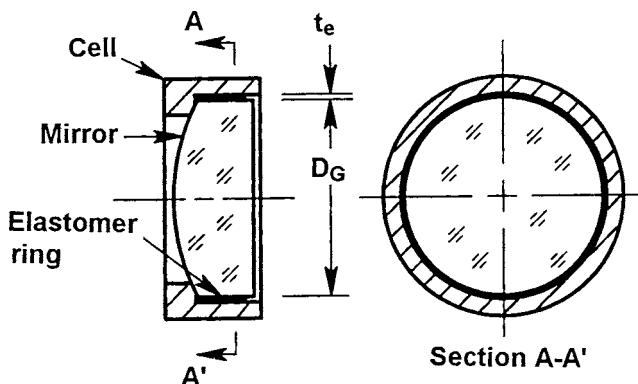
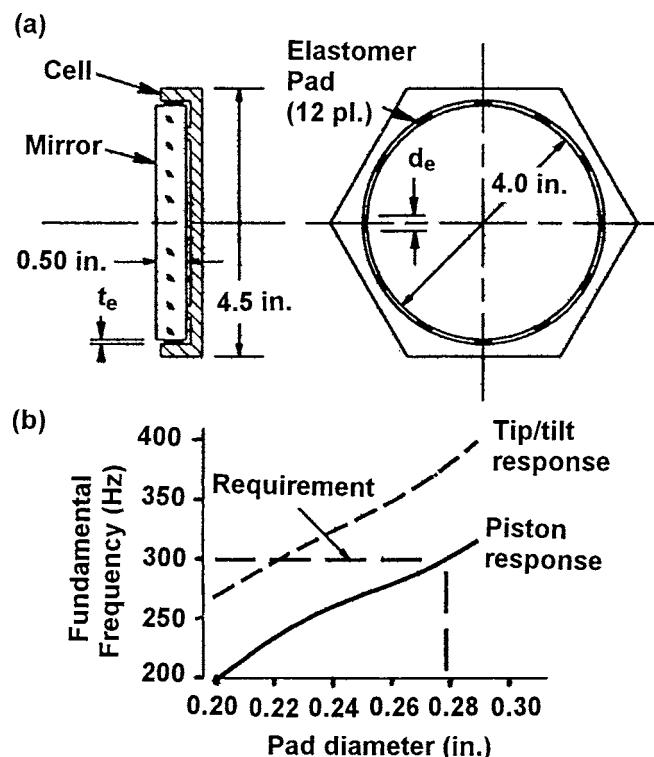


Figure 17.54

(a) A Mirror Mounting Featuring 12 Circular Elastomer Pads Around the Mirror rim Supporting it within the Mount ID
 (b) Graphs of Mirror Subassembly Computed Vibrational Response in Piston and Tip/Tilt for Various Pad Sizes. The System Requirement is also Shown
 (Adapted from Maminini et al.³²)

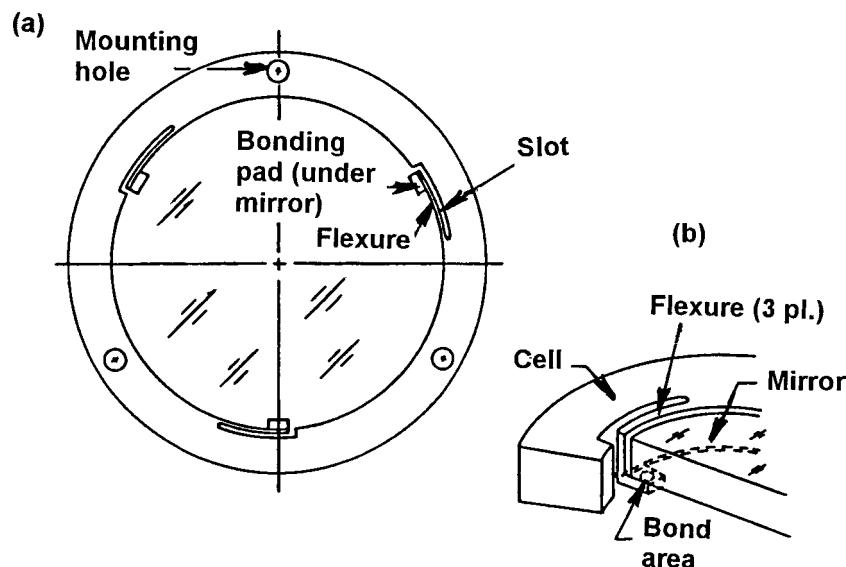


least 0.28 in. (7.1 mm) diameters to raise both these frequencies above the requirement of 300 Hz.

FLEXURE MOUNTINGS A concept for mounting a small circular mirror with cantilevered tangential flexures is depicted in Fig. 17.55. Closely related to the lens mounting concept of Fig. 17.23, this mounting uses flexures that are integral with the body of the ring-shaped mount. Narrow slots made by an electric discharge milling (EDM) process isolate the flexures from the main portion of the mount. These flexures are stiff in the tangential and axial directions and compliant radially as would be appropriate to minimize decentrations caused by temperature changes. A design of this type should be analyzed for vibration response if it is to be used in a dynamic environment.

MOUNTING METAL MIRRORS A preferred method for supporting small metal mirrors involves mounting provisions built into the mirrors themselves. We illustrate a simple case in Fig. 17.56, which shows

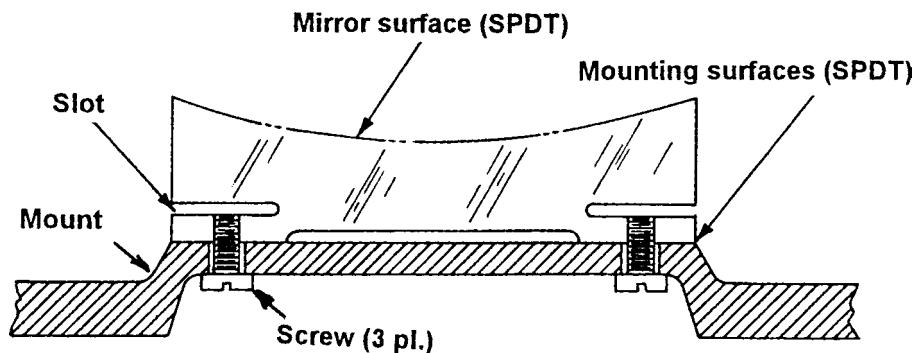
Figure 17.55
Schematic Configuration of a Small Mirror Mounted on 3 Flexures to Ensure Centration Under Temperature Changes (Adapted from Bacich.⁴⁵)



a section through a mirror with three or more machined slots that isolate the mounting “ears” from the main part of the mirror so forces exerted when attaching it to the mount with screws are not transmitted to the optical surface.

A major advantage of metallic mirrors is their compatibility with finish machining by SPDT methods. This process produces precision surfaces with minimal force exerted by the cutting tool on the surface being machined. It also results in accurate relationships between surfaces, especially when they can all be created without removing the part from the machine. When this is not feasible, primary mounting surfaces

Figure 17.56
A Metal Mirror with Diamond-Turned Optical and Mounting Surfaces Constrained by Screws Threaded Directly into Flexure “ears” Machined into the Substrate. (Adapted from Zimmerman.⁴⁶)



can be machined first and then used as the references for turning the optical surfaces and other mounting interfaces.

Figure 17.57a shows an 18 cm diameter symmetric concave aluminum mirror with optical and axial and radial mounting surfaces

Figure 17.57

An Aluminum Mirror and its Mounting (a) Section View of the Mirror (b) Bottom View of the Mirror and (c) Section View of the Mounted Mirror (Adapted from Vukobratovich et al.⁴⁷)

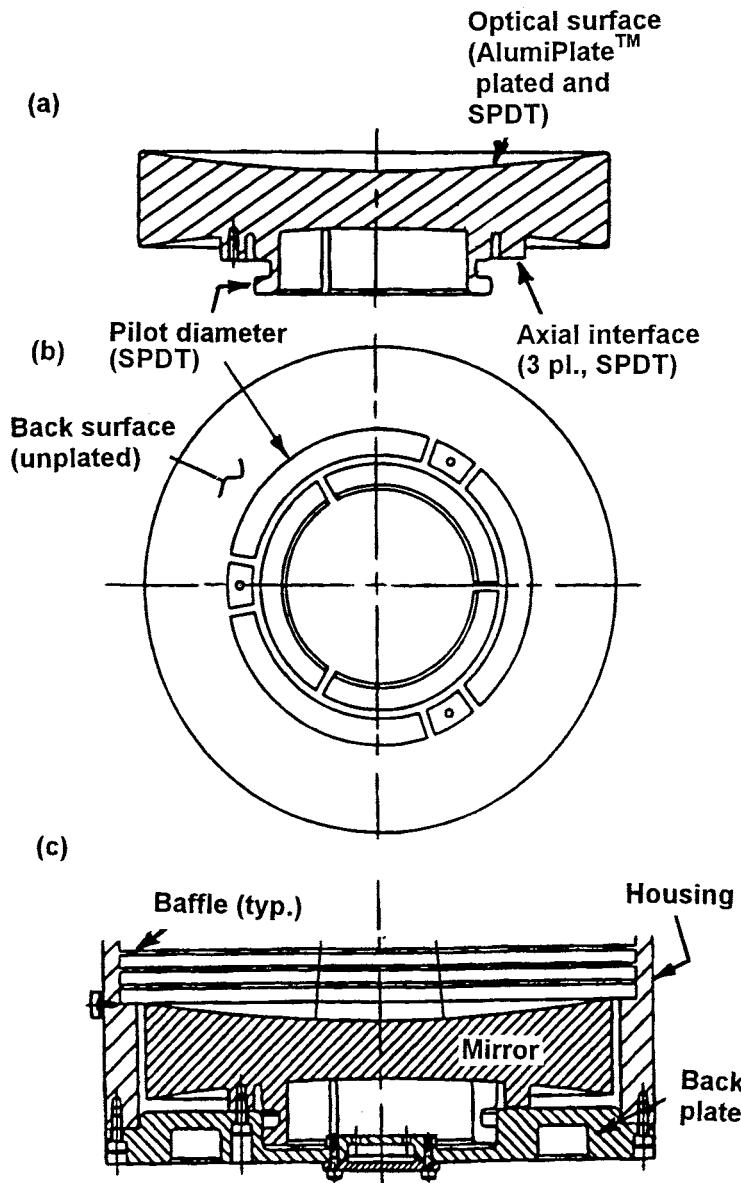
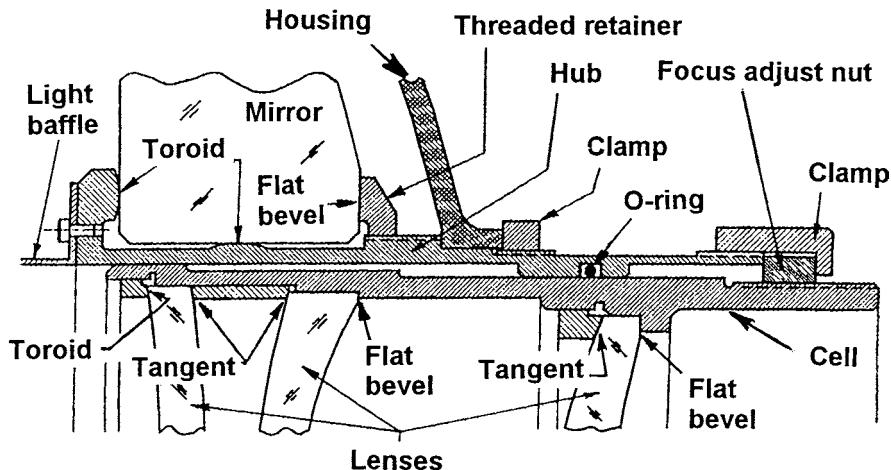


Figure 17.58

A Hub Mounting for a Moderate-Sized Nonmetallic Telescope Primary Mirror. Interfaces with Associated Lenses also are Shown (Adapted from Yoder.²⁵)



diamond turned. The latter surfaces align themselves to diamond turned interfaces on the mount as shown in view (b). The front surface of the mirror was plated with an AlumiPlate coating before finishing. Details of the design were given by Vukobratovich et al.⁴⁷

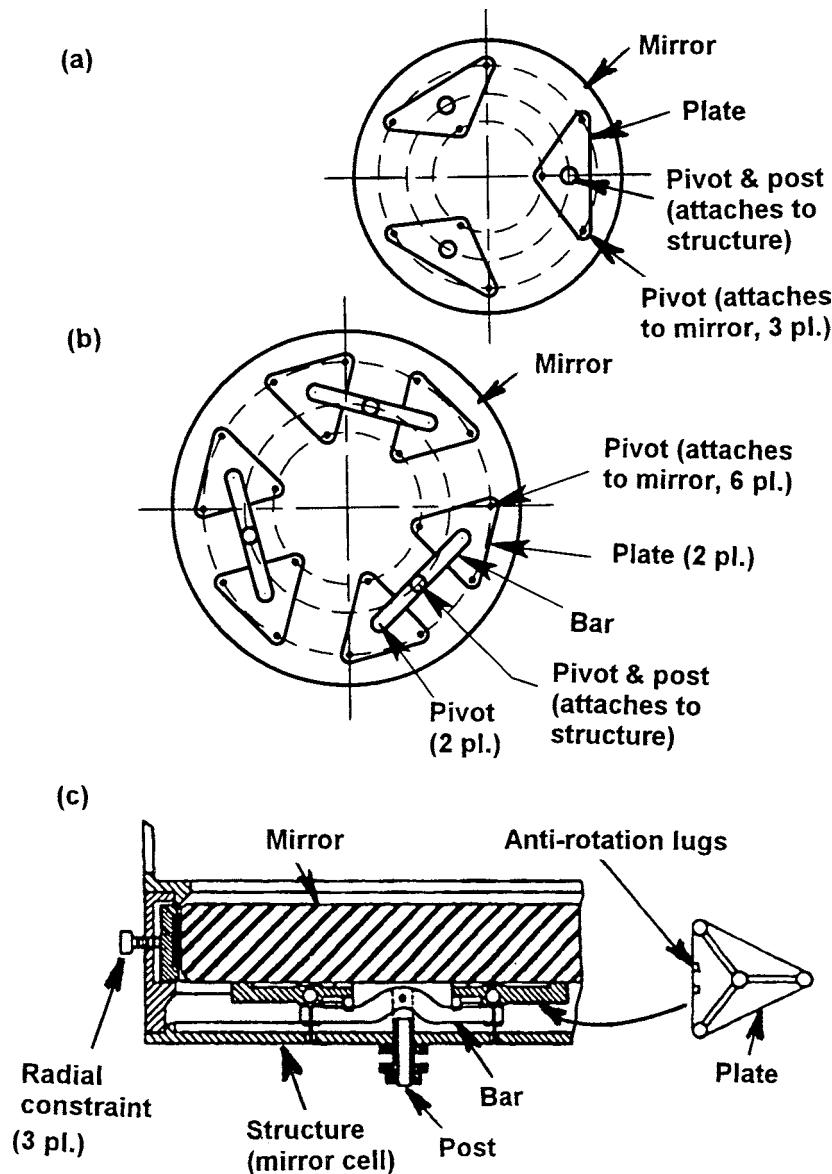
MOUNTING LARGER NONMETALLIC MIRRORS Some moderate-sized nonmetallic mirrors are mounted on a hub that protrudes through a central perforation in the mirror substrate. An example is shown in Fig. 17.58. This is part of the back end of a 3.81 m focal length, *f*/10 catadioptric objective. The first-surface spherical mirror registers against a convex toroidal seat on an integral shoulder of the hub. A toroidal land is provided on the cylindrical hub. The OD of this land is lapped to closely match the ID of the hole in the mirror. A threaded retainer bearing against the flat bevel at the back of the mirror provides the required axial preload.

Mechanical interfaces with three lenses also are indicated in Fig. 17.58. Axial location of the cell containing these lenses is adjusted with the focus adjust nut and clamped. An O-ring seal is provided.

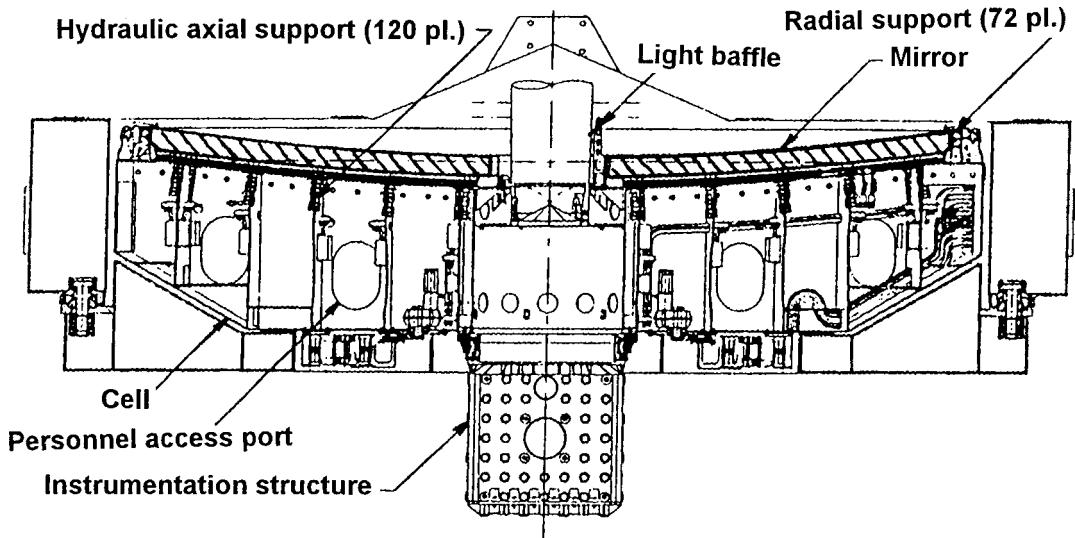
Nonmetallic mirrors at least as large as 2.7 m diameter have been mounted on multiple axial supports as illustrated in Figs. 17.59a and b. These are classical Hindle mounts.⁴⁸ Designs with 9 and 18 supports are shown. The lever mechanism supporting the triangular plates is called a "whiffletree." See view (c) of the figure.

Monolithic cast mirrors to 8.4 m in diameter have been made for use in ground based astronomical telescopes. In spite of their large size, the

Figure 17.59
 Multipoint Mechanical (Hindle) Mounting Systems for Mirrors
 (a) 9-Point Mount (b)
 18-Point Mount (c)
 Side View of one
 Whiffletree Shown in
 (b) (Adapted from
 Hindle.⁴⁸)



substrates are flexible enough that numerous actuators are used to support them axially. The optical figure can then be manipulated under computer control to optimize performance under varying orientations relative to gravity. One such mounting is illustrated in Fig. 17.60. This is

**Figure 17.60**

Schematic Diagram of One Gemini Telescope Primary Mirror and its Supporting Mechanisms (Adapted from Stepp et al.⁴⁹)

for the 8.1 m diameter ultra-low expansion (ULE) primary for one of the Gemini telescopes. Its diameter-to-thickness ratio is 36.8. Most of the mirror's weight is supported when looking vertically by pneumatic pressure over the entire mirror back. A series of 120 combined hydraulic/pneumatic actuators support the balance. A very stiff cell anchors these axial actuators; 72 hydraulic actuators provide radial support.⁴⁹

Mechanical Athermalization Techniques

There are several reasons why performance of an optical system may degrade when the temperature changes: radii, component thicknesses, and refractive indices all change while mechanical component dimensional changes affect air spacings, optical component alignment, and the relative location of the image and sensor. An athermalized optomechanical design has minimized performance variation with temperature. By careful choice of materials and distribution of optical powers, temperature effects on the optical system considered by itself can be reduced. When

the effects of temperature changes on the mechanical system are minimized or that mechanical system is designed to compensate for optical system changes, a higher degree of athermalization can be achieved. Another form of athermalization creates an optomechanical design that maintains preloads over a range of temperatures and hence preserves alignment. In this section, we describe ways in which the mechanical system can be designed to help in achieving athermalization.

"SAME-MATERIAL" DESIGNS Reflecting systems made entirely of the same material scale in dimensions as the temperature changes, but remain in focus and maintain optical performance. An example is shown in Fig. 17.61. This 20 cm aperture Cassegrain telescope is constructed entirely from 6061 aluminum and operates equally from room temperature to 77 K. All optical surfaces and mechanical interfaces are SPDT-machined to ensure alignment when assembled. In addition,

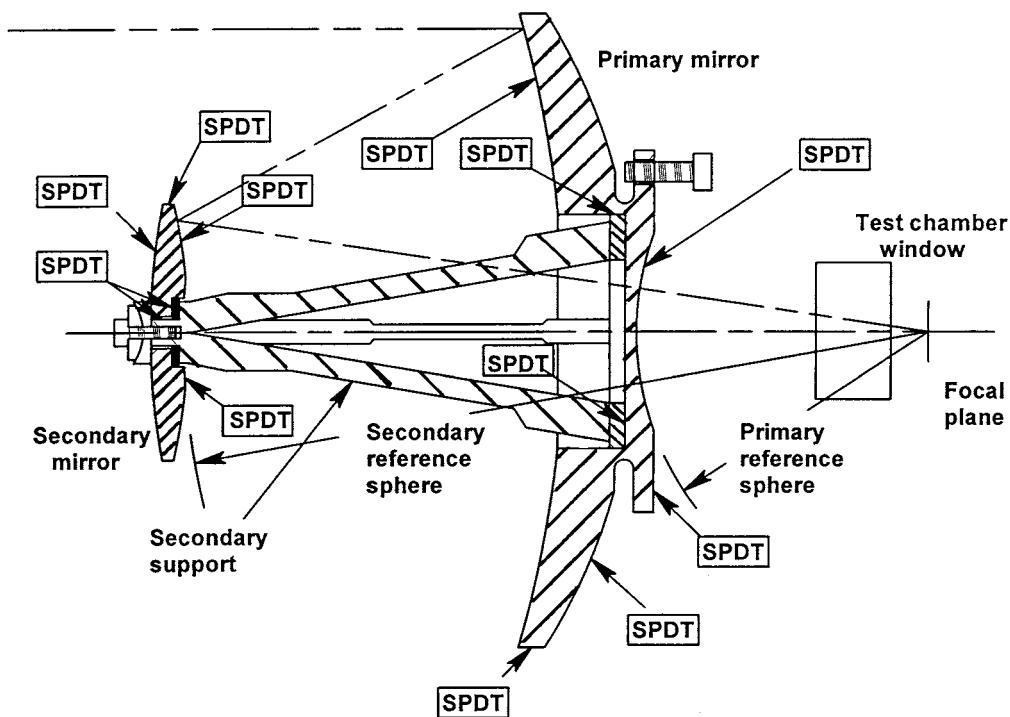


Figure 17.61

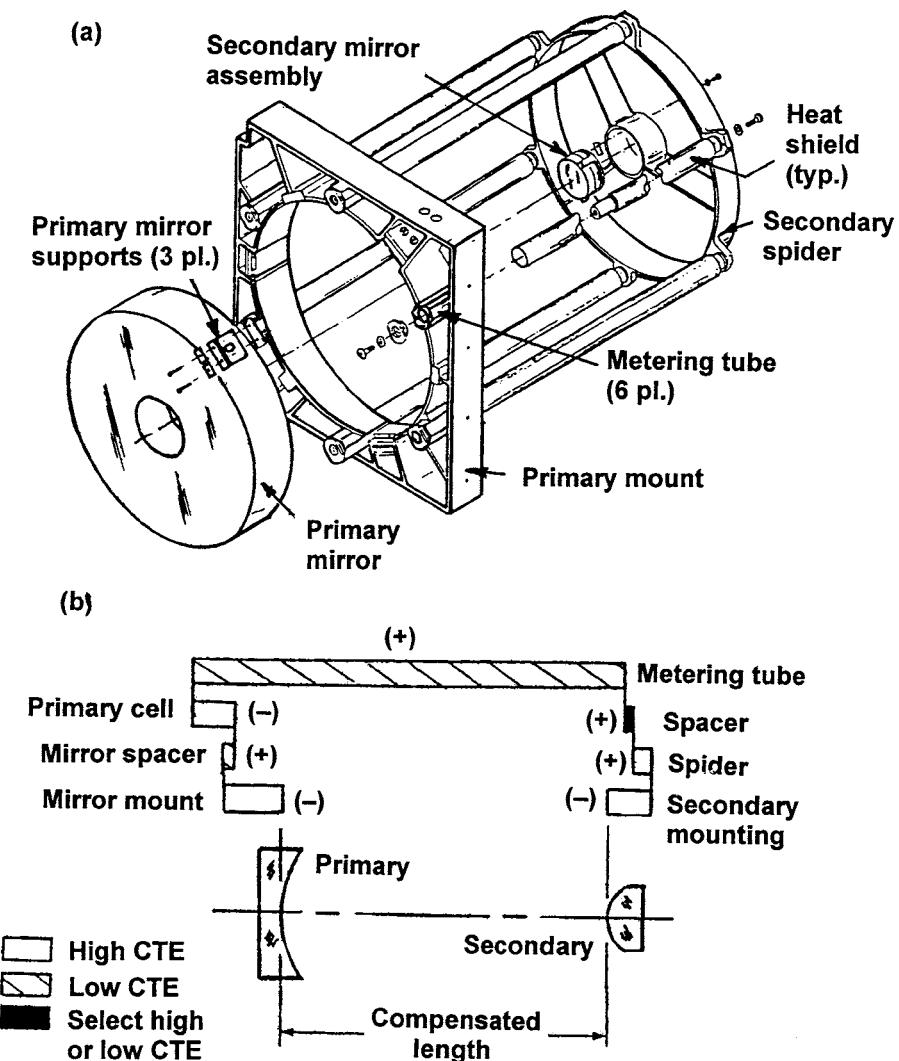
Athermal Cassegrain Telescope Made Entirely from the Same Metal (6061 Aluminum) Optical, Mounting, and Reference Surfaces Marked SPDT are Diamond Turned (Adapted from Erickson et al.⁵⁰)

optical reference surfaces are provided on the mirrors to allow interferometric alignment verification from the image plane. No adjustments are provided nor needed.⁵⁰

PASSIVE ATHERMALIZATION WITH DISSIMILAR MATERIALS

Figure 17.62a shows a 31.1 cm aperture Cassegrain telescope with ULE mirrors that was athermalized for focus by use of dissimilar mechanical materials over the temperature range of 1°C to 54°C. Both mirrors were

Figure 17.62
(a) A Cassegrain Telescope Athermalized with Controlled Lengths of Dissimilar Materials (b) Model of Athermalization Design (Adapted from Zurnehly and Hookman.⁵¹)



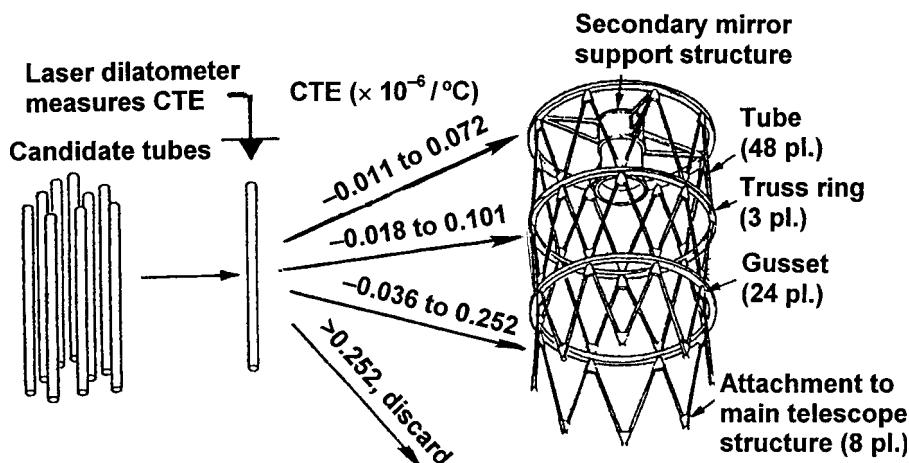
mounted in aluminum mounts separated by 6 Invar tubes. Spacers were customized at assembly to establish initial focus.

Figure 17.62b depicts the concept for athermalization of this telescope. The bars represent different lengths of metals with low CTE (diagonally shaded) or moderate CTE (unshaded). The algebraic sign associated with each bar indicates how a positive temperature change of that part affects the mirror separation. The axial lengths of the parts are chosen so that separation remains constant. The material for the secondary spacer is chosen at time of assembly as low or moderate CTE to compensate for dimensional length variations from nominal of other components.⁵¹

The separation between the primary and secondary mirrors of the Hubble Space Telescope was stabilized in a similar manner. The structure connecting the mirror mounts was designed as a 4.9 m long, 3-bay truss made of tubes, rings, and gussets. Low CTE metals, such as Invar, could not be used because of severe weight constraints so the structure was constructed of graphite epoxy. The truss was designed to hold mirror despace to $\pm 1.5 \mu\text{m}$, decenter to $10 \mu\text{m}$, and tilt to 2 arcsec in spite of orbital temperature excursions of $\sim 30^\circ\text{C}$. To achieve this, the effective CTE of the complete truss would need to be nearly zero.

To compensate for inevitable manufacturing variability, the actual CTEs of the truss tubes were measured and those tubes were used at different locations in the truss as indicated in Fig. 17.63. For example, ones with the highest acceptable CTE variability were used in the bay nearest the primary mirror where the operational temperature variation

Figure 17.63
Mechanically
Athermalized
Graphite/Epoxy Truss
that Supports the
Secondary Mirror at
Constant Distance
from the Primary in
the Hubble Space
Telescope. (Adapted
from McCarthy and
Facey⁵²)

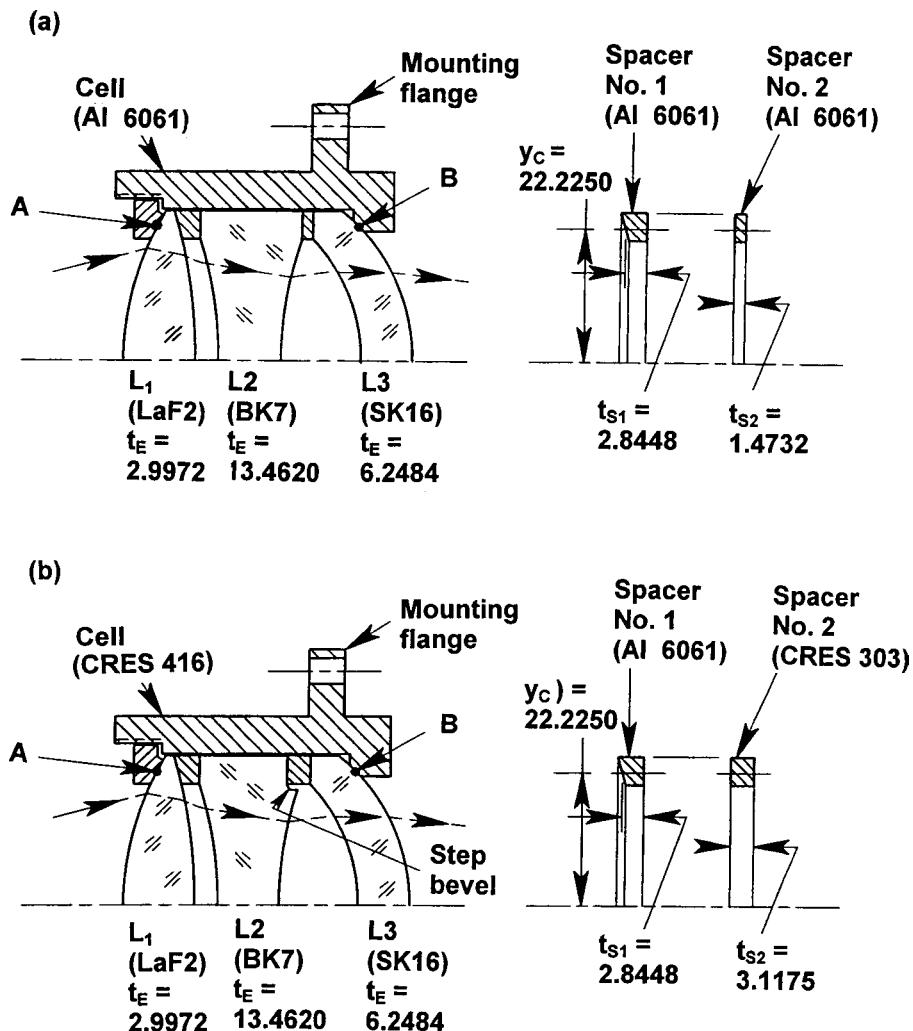


would be the least. The effects of some tubes with negative CTEs were balanced with those of tubes with positive CTEs.⁵² The success of both the design and hardware manufacture is evidenced by the outstanding imagery produced by the Hubble telescope during long exposures.

A refracting lens assembly design that could be improved by mechanical athermalization with dissimilar materials is shown in Fig. 17.64a. All metal parts of this assembly are made of 6061 aluminum while the lenses are made of the indicated glasses. Analysis indicated that, at maximum

Figure 17.64

(a) An Air Spaced Triplet Lens Assembly that Is not Mechanically Athermal so may Suffer Misalignment at High Temperature
 (b) Modified Design Athermalized Mechanically by use of Some Metals with Lower CTEs and a Thicker Spacer
 (Adapted from Yoder and Hatheway.³⁸)



temperature, expansion of the cell over the 27.0260 mm length from point "A" to point "B" (at the height of contact y_d) would exceed that of the lenses and spacers sufficiently for the axial preload applied at assembly to be dissipated. The lenses would then be free to move under vibration and misalignment could result.

To prevent this from happening with simple changes in components, metals with different CTEs could be employed as indicated in Fig. 17.64b. A step bevel would be added to L_2 to allow the second spacer to be lengthened so the design would become axially athermal from "A" to "B." Then, temperature changes would not significantly affect preload or optical alignment.¹⁰ Note that these changes do not ensure good optical performance nor constant focus at all temperatures. Those aspects of the design need to be taken care of as well.

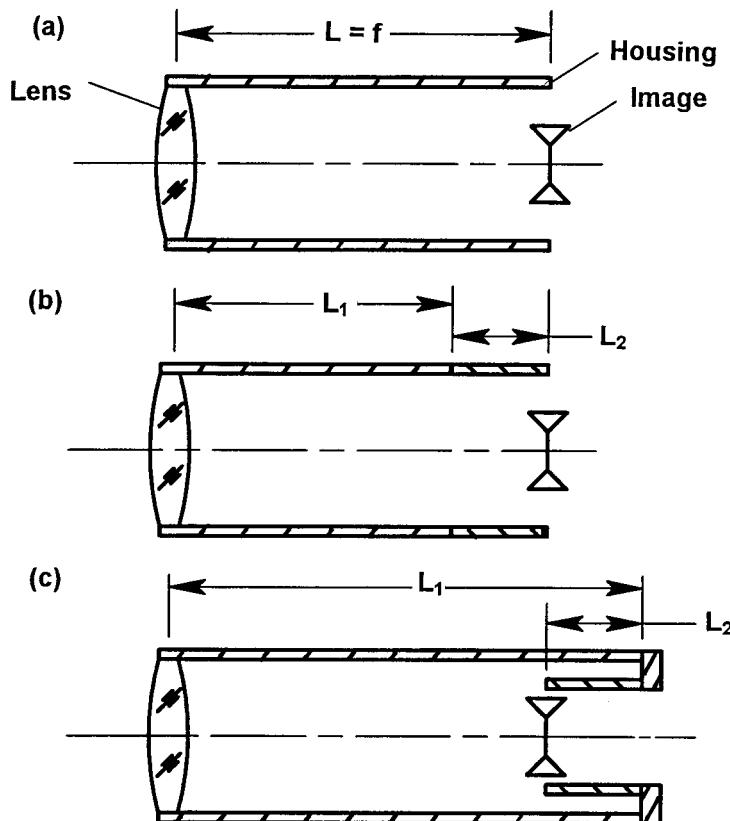
A general technique for maintaining focus of a refracting system as the temperature changes is as follows. With a lens design program, the lens designer computes the change in location of the *best* image with temperature. The mechanical designer then creates a mechanical structure that moves the sensor as required to closely approximate that image location at all temperatures. Figure 17.65 shows an elementary example. In view (a), a single lens forms its image on a sensor (such as film) at the end of a tubular metallic housing. Defocus results when the temperature changes. To compensate, mechanical housings comprising specific lengths of two materials with different CTEs are employed. In view (b), the tube lengths expand in the same direction with increasing temperature while, in view (c), they expand in opposite directions. The latter case is called a reentrant design. In both cases, the lengths L_1 and L_2 are chosen so the total change in mechanical length equals the change in system back focal distance. Athermalization of focus would then be achieved over the temperature range of constancy of CTEs.

ACTIVE CONTROL OF FOCUS One technique for athermalizing focus of optical systems is active control of the axial location(s) of one or more optic in which the temperature distribution within the system is measured and motor-driven mechanisms are used to drive the optic separation and/or final image distance to optimum values in accordance with preestablished algorithms.

An example of such a compensation technique is the 5:1 afocal zoom attachment for a military forward looking infrared (FLIR) sensor operating in the spectral range from 8 to 12 μm . The optomechanical design is shown in Figs. 17.66a and b. The first element is fixed, as are the smaller

Figure 17.65

Athermalization of a Simple Refracting System (a) with Differing Lengths of Structural Tubes having Different CTEs: (b) Series Configuration (c) Re-entrant Configuration

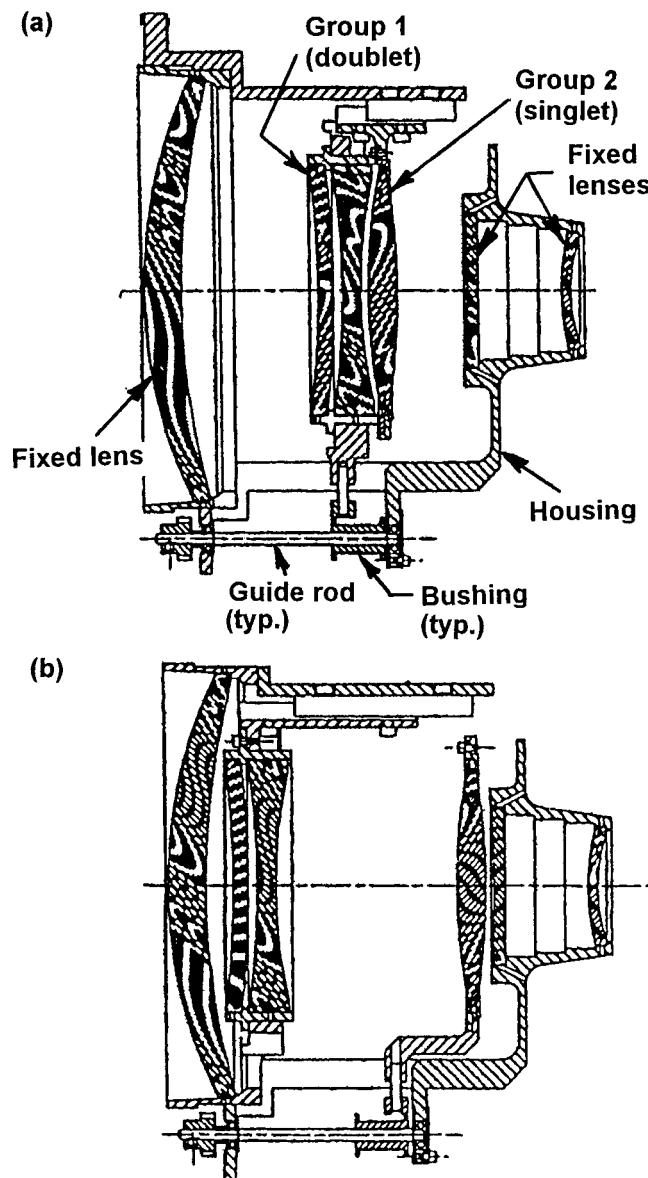


lenses at right. The moveable lenses are designated Groups 1 (air-spaced doublet) and 2 (singlet). All of these lenses are made of germanium, as is the second small fixed lens. The other small fixed lens is zinc selenide. There are four aspherics in the design. Image quality of this design would meet all requirements over the specified temperature and target distance ranges if the locations of the moveable lens groups could be reoptimized for each combination of operating parameters. The usual technique of driving the lens motions by one or two mechanical cam(s) will not suffice here because there are too many variables.

To athermalize this design, the moveable lens groups are attached to linear bushings that slide on guide rods. Two stepper motors acting through appropriate gear trains drive them independently. See Fig. 17.67a. The motors are controlled during operation by a local microprocessor. The operator commands the magnification to be provided

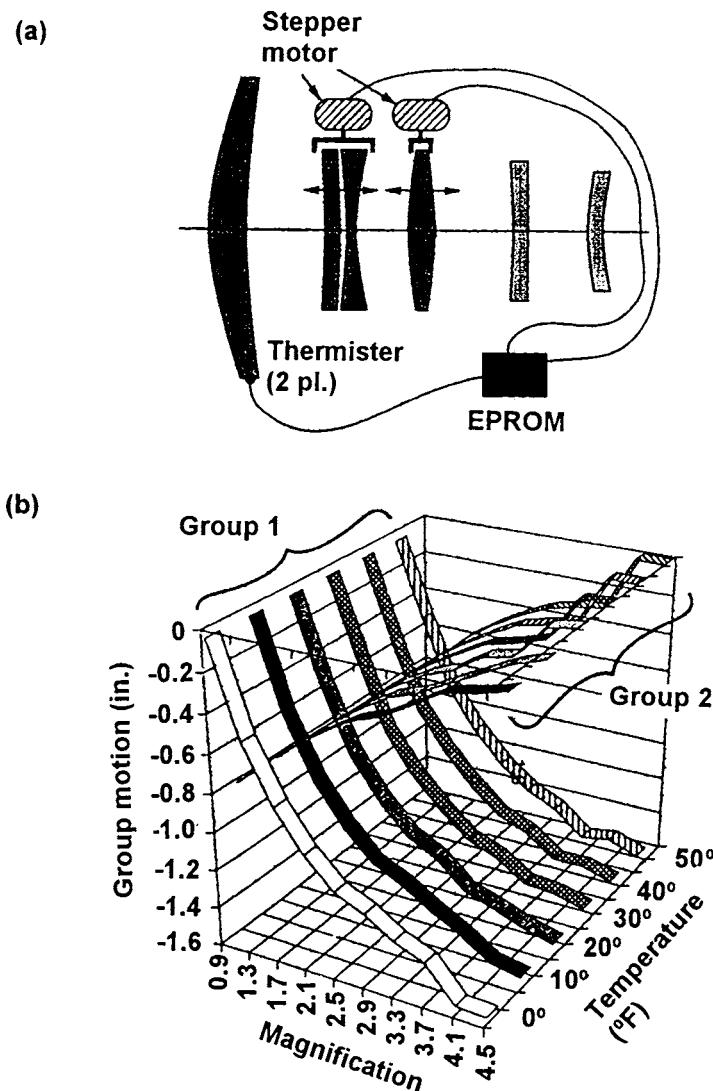
Figure 17.66

Optomechanical System for an Athermalized Zoom Lens (a) at Telephoto (High Magnification) Setting (b) at Wide Angle (Low Magnification) Setting (Adapted from Fischer and Kampe.⁵³)



and the target range. The electronics system then refers to a look-up table stored in a built-in *erasable programmable read only memory* (EPROM) to determine the appropriate settings for the moveable lenses at room temperature. Thermistors attached to the lens housing sense the temperature of the assembly. Signals from these sensors are used

Figure 17.67
 (a) Lens Motion Control System for the Assembly of Figure 17.66. (b) Lens Group Motions as Functions of Magnification and Temperature at Constant Target Range (Adapted from Fischer and Kampe.⁵³)



by the electronics to select, from a second look-up table stored in the EPROM, the required refinements of the lens settings to correct for temperature effects on system focus. The corrected signals then drive the motors to position the lenses for best imagery at the measured temperature. The lens group motions vary as functions of magnification and temperature as indicated in Fig. 17.67 b. Similar relationships exist for group motion variations as functions of magnification and target range at constant temperature.⁵³

References

1. ASTM E595-93(2003)e2, *Standard Test Method for Total Mass Loss and Collected Volatile Condensable Materials from Outgassing in a Vacuum Environment*, ASTM International, West Conshohocken, PA.
2. Tribble, A. C., *Fundamentals of Contamination Control*, TT44, SPIE Press, 2000.
3. MIL-STD-210, *Climatic Information to Determine Design and Test Requirements for Military Systems and Equipment*. U.S. Dept. of Defense, Washington.
4. ISO Specification 10109, *Optics and Optical Instruments—Environmental Requirements*.
5. Tribble, A. C., *The Space Environment*, Princeton University Press, Princeton, 1995.
6. Sarafin, T. P., ed., *Spacecraft Structures and Mechanisms*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.
7. Shipley, A. F., "Optomechanics for Space Applications," *SPIE Short Course SC561*, 2007.
8. U.S. Military Specification MIL-STD-810, "Environmental Engineering Considerations and Laboratory Tests."
9. ISO Specification 9022, "Environmental Test Methods" (22 Parts).
10. Yoder, P. R., Jr., *Opto-Mechanical Systems Design*, 3d. ed., CRC Press, Boca Raton, 2005.
11. Paquin, R. A., "Materials for Optical Systems," Chapter 3 in *Handbook of Optical Engineering*, A. Ahmad, ed., CRC Press, Boca Raton, 1997.
12. Vukobratovich, D., "Optomechanical Design Principles," Chapter 2 in *Handbook of Optical Engineering*, A. Ahmad, ed., CRC Press, Boca Raton, 1997.
13. Yoder, P. R., Jr., "Two New Lightweight Military Binoculars," *J Opt Soc Am*, 50, 1960: 491.
14. Trsar, W. J., Benjamin, R. J., and Casper, J. F., "Production Engineering and Implementation of a Modular Military Binocular," *Opt. Eng.* 20, 1981: 201.
15. Visser, H. and Smorenburg, C., "All Reflective Spectrometer Design for Infrared Space Observatory," *Proc. SPIE*, 1113, 1989: 65.
16. Krim, M., "Mechanical Design of Typical Systems for Space Operation," *Proc. SPIE*, CR43, 1992: 3.
17. Paquin, R. A., "Metal Mirrors," Chapter 4 in *Handbook of Optomechanical Engineering*, A. Ahmad, ed., CRC Press, Boca Raton, 1997.

18. Vukobratovich, D., "Lightweight Mirror Design," Chapter 5 in *Handbook of Optomechanical Engineering*, A. Ahmad, ed., CRC Press, Boca Raton, 1997.
19. Serrurier, M., "Structural Features of the 200-Inch Telescope for Mt. Palomar Observatory," *Civil Eng.*, 8, 1938.
20. Hopkins, R. E., "Lens Mounting and Centering," Chapter 2 in *Applied Optics and Optical Engineering*, Vol. VIII, R. R. Shannon and J. C. Wyant, eds., Academic Press, New York, 1980.
21. Karow, H. H., *Fabrication Methods for Precision Optics*, Wiley, New York, 1993.
22. Ford, V. G., et al., "Optomechanical Design of Nine Cameras for the Earth Observing System Multi-Angle Imaging Spectro-Radiometer, TERRA Platform," *Proc. SPIE*, 3786, 1999: 264.
23. Barkhouser, R. H., Smee, S. A., and Meixner, M., "Optical and Optomechanical Design of the WIYN High Resolution Infrared Camera," *Proc. SPIE*, 5492, 2004: 921.
24. Barrera, S., et al., "EMIR Optomechanics," *Proc. SPIE*, 5495, 2004: 611.
25. Yoder, P. R., Jr., *Mounting Optics in Optical Instruments*, 2d. ed., SPIE Press, Bellingham, 2008.
26. ASME Publication B1.1—2003, *Unified Inch Screw Threads, UN and UNR Thread Form*, ASME, New York.
27. Roark, R. J., *Formulas for Stress and Strain*, 3d. ed., McGraw-Hill, New York, 1954.
28. Bayar, M., "Lens Barrel Optomechanical Design Principles," *Opt. Eng.*, 20, 1981: 181.
29. Herbert, J. J., "Techniques for Deriving Optimal Bondlines for Athermal Bonded Mounts," *Proc. SPIE*, 6288, 2006.
30. Doyle, K. B., Michels, G. J., and Genberg, V. L., "Athermal Design of Nearly Incompressible Bonds," *Proc. SPIE*, 4771, 2002: 296.
31. Michels, G. J., Genberg, V. L., and Doyle, K. B., "Finite Element Modeling of Nearly Incompressible Bonds," *Proc. SPIE*, 4771, 2002: 287.
32. Mammini, P., et al., "Sensitivity Evaluation of Mounting Optics using Elastomer and Bipod Flexures," *Proc. SPIE*, 5176, 2003: 26.
33. Bruning, J. H., DeWitt, F. A., and Hanford, K. E., "Decoupled Mount for Optical Element and Stacked Annuli Assembly," *U. S. Patent No. 5,428,482*, 1995.
34. Delgado, R. F. and Hallinan, M., "Mounting of Lens Elements," *Opt. Eng.*, 14, 1975: S-11. (Reprinted in *SPIE Milestone Series*, 770, 1988: 173.)
35. Harris, D. C., *Materials for Infrared Windows and Domes*, SPIE Press, Bellingham, 1999.

36. Vukobratovich, D., "Introduction to Opto-Mechanical Design," *SPIE Short Course SC014*, 2003.
37. Timoshenko, S. P. and Goodier, J. N., *Theory of Elasticity*, 3d. ed., McGraw-Hill, New York, 1970.
38. Yoder, P. R., Jr. and Hatheway, A. E., "Further Considerations of Axial Preload Variations with Temperature and the Resultant Effects on Contact Stresses in Simple Lens Mountings," *Proc. SPIE*, 5877, 2005: 587705.
39. Quammen, M. L., et al., U.S. Patent 3,246,563, 1966.
40. Cassidy, L. W., "Advanced stellar sensors—a new generation," in: *Technology for Space Astrophysics Conference: The Next 30 Years*, Digest of Papers, AIAA/SPIE/OSA Symposium, Danbury, CT, 1982: 164—173.
41. MIL-HDBK-141, *Optical Design*, Defense Supply Agency, Washington, 1962.
42. Fischer, R. E., "Case Study of Elastomeric Lens Mounts," *Proc. SPIE*, 1533, 1991: 27.
43. Williamson, D. M., "Compensator Selection in the Tolerancing of a Microlithography Lens," *Proc. SPIE*, 1049, 1989: 178.
44. Durie, D. S. L., "Stability of Optical Mounts," *Machine Design*, 40, 1968: 184.
45. Bacich, J. J., "Precision lens mounting," *U.S. Patent No. 4,733,945*, 1988.
46. Zimmerman, J., "Strain-Free Mounting Techniques for Metal Mirrors," *Opt. Eng.*, 20, 1981: 187.
47. Vukobratovich, D., Gerzoff, A., and Cho, M. K., "Therm-Optic Analysis of Bi-Metallic Mirrors," *Proc. SPIE*, 3132, 1997: 12.
48. Hindle, J. H., "Mechanical Flotation of Mirrors," *Amateur Telescope Making, Book One* (A. G. Ingalls, ed.), Scientific American, New York, 1945.
49. Stepp, L., Huang, E., and Cho, M., "Gemini Primary Mirror Support System," *Proc. SPIE*, 2199, 1994: 223.
50. Erickson, D. J., Johnston, R. A., and Hull, A. B., "Optimization of the Optomechanical Interface Employing Diamond Machining in a Concurrent Engineering Environment," *Proc SPIE*, CR43, 1992: 329.
51. Zurmehly, G. E. and Hookman, R., "Thermal/Optical Test Setup for the Geostationary Operational Environmental Satellite Telescope," *Proc. SPIE*, 1167, 1989: 360.
52. McCarthy, D. J. and Facey, T. A., "Design and Fabrication of the NASA 2.4-Meter Space Telescope," *Proc. SPIE*, 330, 1982: 139.
53. Fischer, R. E. and Kampe, T. U., "Actively Controlled 5:1 Afocal Zoom Attachment for Common Module FLIR," *Proc. SPIE*, 1690, 1992: 137.

Optical Manufacturing Considerations

From the point of view of a lens manufacturer, what design attributes have the most influence on manufacturing efficiency? The primary design considerations are optical material, component size, shape, and manufacturing tolerances. All of these attributes are variable at the design phase and can have significant impact on lens manufacturing costs.

In order to narrow the scope of this chapter, the text assumes the manufacture of a precision glass lens of approximately 50-mm diameter using grinding and polishing techniques. The information is presented in the following order:

1. *Material.* A summary of manufacturing considerations for optical glasses
2. *Manufacturing.* An overview of conventional and advanced process technologies
3. *Special fabrication considerations.* A review of tolerancing trade-offs and finishing options
4. *Relative manufacturing cost.* An analysis of manufacturing variables
5. *Sourcing considerations.* Suggestions for achieving project goals
6. *Conclusion.* A summary table for quick reference

While this analysis is based on a 50-mm-diameter glass lens, it can also be adapted to include specific market niches such as microoptics (diameters smaller than 5 mm), macrooptics (diameters larger than 300 mm), prisms and flats, molded glass and plastic optics, diamond-turned crystal and metal optics, and diffractive optical elements. These niches are addressed in additional chapters of this book.

Material

There are more than 100 different optical glasses available worldwide, and each has a unique set of optical, chemical, and thermal characteristics. Only a few glass manufacturers in the world produce these optical glasses, and each manufacturer has a company-specific glass-naming convention. Cross-referencing the glasses is possible via a six-digit glass code (ABCXYZ) that is derived from the index of refraction ($n_d = 1.ABC$) and the Abbe value ($v_d = XYZ$). For the vast majority of optical applications, glasses from differing manufacturers can be direct substitutes. Lens designers should be aware, however, that equivalent glasses having the same six-digit glass code might not have exactly the same optical, chemical, and mechanical properties. For example, Schott's SK-16 (620603) has slightly different characteristics than Ohara's S-BSM-16 (620603). Be aware that optical design software will define glasses that can achieve a desired optical performance, but it cannot determine the glasses' current availability in the market. Nor will the software give consideration for the glasses' chemical and thermal properties. For example, it may be important to consider that the index of refraction of a glass changes with temperature at a known rate. Other parameters that are important to consider are spectral transmission, dispersion, material quality, and mechanical, chemical, and thermal properties.

Design Considerations

Material quality is defined by tolerances of optical properties, striae grades, homogeneity, and birefringence. Optical properties include spectral transmission, index of refraction, and dispersion. Data for each glass type is available from its manufacturer. If tighter than standard optical properties are required, then additional cost and time are

usually associated with obtaining the material. Specification of glass based on material quality is provided in the International Standard ISO 10110 and the U.S. military specification MIL-G-174B. A brief summary of glass material specifications using nomenclature from Schott optical glass is shown in Fig. 18.1.

Before finalizing an optical design, some consideration should be given to glass cost and availability. Glass prices vary from a few dollars per pound to several hundred dollars per pound. In some cases, it may be more economical to add a lens to the design in order to avoid expensive glasses. In addition, many glasses are not regularly stocked; instead they are melted to order, which can take several months. Pricing and melt frequencies are available from glass manufacturers. Each manufacturer has a

Figure 18.1
Glass Material
Specifications

Striae Grade AA (P) is classified as “precision striae” and has no visible striae. Grade A only has striae that are light and scattered when viewed in the direction of maximum visibility. Grade B has only striae that are light when viewed in direction of maximum visibility and parallel to the face of the plate.

Birefringence is the amount of residual stress in the glass and depends on annealing conditions, type of glass, and dimensions. The birefringence is stated as nanometers per centimeter difference in optical path measured at a distance from the edge equaling 5% of the diameter or width of the blank. Normal quality is defined as (except for diameters larger than 600 mm and thicker than 100 mm):

1. Standard is less than or equal to 10 nm/cm.
2. Special annealing (NSK) or precision annealing is less than or equal to 6 nm/cm.
3. Special annealing (NSSK) or precision quality after special annealing (PSSK) is less than or equal to 4 nm/cm.

Homogeneity is the degree to which refractive index varies within a piece of glass. The smaller the variation, the better the homogeneity. Each block of glass is tested for homogeneity grade.

Normal grade	$\pm 1 \times 10^{-4}$
H1 grade	$\pm 2 \times 10^{-5}$
H2 grade	$\pm 5 \times 10^{-6}$
H3 grade	$\pm 2 \times 10^{-6}$
H4 grade	$\pm 1 \times 10^{-6}$

Tolerances of optical properties consist of deviations of refractive index for a melt from values stated in the catalog. Normal tolerance is ± 0.001 for most glass types. Glasses with n_d greater than 1.83 may vary by as much as ± 0.002 from catalog values. Tolerances for n_d are ± 0.0002 for grade 1, ± 0.0003 for grade 2, and ± 0.0005 for grade 3.

The dispersion of a melt may vary from catalog values by $\pm 0.8\%$. Tolerances for v_d are $\pm 0.2\%$ for grade 1, $\pm 0.3\%$ for grade 2, and $\pm 0.5\%$ for grade 3.

list of “preferred” glasses that are most frequently melted and usually available from stock. It’s important to note that preferred does not imply “best glass type available.” From a manufacturing perspective, preferred refers only to the availability of the glass in stock. For example, BK-7 is readily available from stock and is among the most economical of glass types. On the other hand, a glass like SF-59 is not made as frequently and may not be as readily available. If delivery is a concern, the designer may want to use only glasses from the frequently melted glass list.

Fabrication Considerations

Since the mechanical, chemical, and thermal properties of glass are what determine the ease or difficulty of making optics from the material, these properties are of particular interest to the optical fabricator.

MECHANICAL PROPERTIES *Mechanical properties* include hardness and abrasion resistance. These properties determine the rate at which material is removed, and should be among the first to consider.

Hardness is measured in accordance with International Standards Organization (ISO) 9385. It is measured with a microhardness tester that utilizes a precision diamond point applied with a specific amount of force. This probe contacts and penetrates the polished glass sample at room temperature. Carefully measuring the resultant indentation yields a calculation known as the “Knoop hardness” of the material. Knoop hardness ranges from 300 to 700 for most optical glasses, where 300 represents a soft glass and 700 harder glasses. In general, the harder the glass, the longer the time required to grind and polish the lens.

Abrasion resistance describes how fast the glass will process. Abrasion resistance is the ratio of material removed on a test piece of glass to the material removed from a BK-7 sample. The abrasion resistance of BK-7 is set to equal 100. The higher the number, the faster the material will be removed. The values range from about 60 to 400. Compared to BK-7, a glass with a value of 60 will take almost twice as long to process. Conversely, glass with a value of 400 will take only one-quarter of the time. The process time seems to imply that softer glasses are cheaper to fabricate. One must remember, however, that other factors, such as cosmetic finish, may offset potential savings. Soft glasses are more difficult to polish to achieve very good cosmetics and low rms surface roughness. As a general rule of thumb, for lenses with identical specifications, except for material,

a BK-7 lens will be cheaper to produce. The cost of a lens increases as the abrasion resistance value moves away from that of BK-7. For example, glasses that have high abrasion resistance can require significantly longer grinding and polishing times. On the other hand, glasses with a low abrasion resistance are more difficult to achieve tight thickness tolerance, especially when good cosmetics are required.

RELATIVE COST AND DENSITY *Relative cost and density* are also important factors to consider. The density of glass is described in grams per cubic centimeter. Multiplying this number by the blank volume (including cutting allowances) and cost yields the approximate cost of a blank. It is important to remember dollars per pound of glass is not the only factor that determines the cost of the optic. For example, SF6 and SFL6 are virtually identical optically. SFL6 costs 63% more per pound, but its density is only 65% of the density of SF6, offsetting the higher per pound cost. In addition, SFL6 is much easier to process, which ultimately results in lower manufacturing costs.

CHEMICAL PROPERTIES *Chemical properties* are also of interest to the optician. There are several tests that characterize the chemical behavior of glass with regard to humidity, acid, alkali, and phosphate stainability. The values reported from these tests reflect the degree of processing difficulty and special handling a glass will require. Designers should, therefore, refer to chemical property test values when making lens design decisions. The chemical properties tests for glass are explained in more detail in Fig. 18.2.

To summarize the chemical properties listed in Fig. 18.2, if a glass is low in all categories, then it is stable and unlikely to stain during standard manufacturing processes and storage. If a glass is high in one or more categories, it is very likely to cause problems if special care is not taken. As a general rule, any glass with a stain coefficient of three or more must be handled with special care. Glasses with stain designations in the 50s (for example, SK-55 or S-FPL53) tend to be very troublesome. The poor chemical properties of these glasses can lead to residual stain from deblocking, cleaning, and/or handling of the lens. If stained, the lens may require repolishing to remove the stain. This causes more risk to the part, either from handling or missing the mechanical tolerances. For example, if tight thickness control is required and the glass is prone to staining, it is more difficult to achieve a stain-free surface within the desired thickness tolerance.

Figure 18.2
Chemical Property
Tests for Glass

Climate resistance (CR) is a test that evaluates the material's resistance to water vapor. Glasses are rated and segregated into classes, CR 1 to CR 4. The higher the class, the more likely the material will be affected by high relative humidity. In general, all optically polished surfaces should be properly protected before storing. Class 4 glasses should be processed and handled with extra care.

Resistance to acid (SR) is a test that measures the time taken to dissolve a 0.1- μm layer in an aggressive acidic solution. Classes range from SR 1 to SR 53. Glasses of classes SR 51 to SR 53 are especially susceptible to staining during processing and require special consideration.

Resistance to alkali (AR) is similar to resistance to acid because it also measures the time taken to dissolve a 0.1- μm layer, in this case, in an aggressive alkaline solution. Classes range from SR 1 to SR 4, with SR 4 being most susceptible to stain from exposure to alkalis. This is of particular interest to the optician because most grinding and polishing solutions become increasingly alkaline due to the chemical reaction between the water and the abraded glass particle. For this reason most optical shops monitor the pH of their slurries and adjust them to neutral as needed.

Resistance to staining (FR) is a test that measures the stain resistance to slightly acidic water. The classes range from FR 0 to FR 5, with the higher classes being less resistant. The resultant stain from this type of exposure is a bluish-brown discoloration of the polished surface. FR 5 class lenses need to be processed with particular care since the stain will form in less than 12 min of exposure. Hence, any perspiration or acid condensation must be removed from the polished surface immediately to avoid staining. The surface should be protected from the environment during processing and storage.

THERMAL PROPERTIES *Thermal properties* of glass may also affect optimal process methods. Thermal expansion coefficients range from 4 to $16 \times 10^{-6}/\text{K}$. Glasses with a coefficient over 10 must be handled very carefully during any operation involving rapid thermal change. In fact, even body heat that is transferred by touching the glass may cause subsurface microfractures. Glasses with high thermal coefficients of expansion are more susceptible to surface distortion and catastrophic fractures during blocking and handling. If the coefficient is over 10, then the process should not include any rapid thermal processes. Due to the difficulty in handling these glasses, they should be avoided whenever possible.

Optical glasses can be segregated into groups by their material properties. It may be helpful to contact the preferred glass manufacturer for a particular material to get summary data. As an example, Table 18.1 is a quick reference chart for selecting the more favorable glasses.

TABLE 18.1

Optical Glasses
Categorized by
Material Properties

Stable Glasses	Climate Stainable	Alkaline Stainable	Acid Stainable	Heat Sensitive	Soft Material
BK7	PSK50	LaK11	FK3	FK52	FK3
BaK2	SK16	LaK21	PSK52	PSK53A	PSK54
SFL6	LaK21	All KzFS	SK16	SF59	SF6
SF11	KzFS1	SK16	SSKN5	TiF6	TiF6

Glass material is available in various forms of supply. It can be in block, rod, or slab form, requiring sawing or core drilling operations to make it into disks or it can be purchased as a disk. In any case, the desired form is defined by a diameter and a thickness, or, in other words, a cylinder that totally contains the final lens geometry with some oversize allowance for processing. This approach is the quickest but not the most cost effective. Buying the glass as a molded blank results in the lowest cost. Material efficiency is achieved by taking a piece of glass of the appropriate weight, heating it, and pressing it into a metal mold to make the shape (slightly larger) of the final lens. This approach requires several weeks for the glass to be delivered; however, it minimizes glass cost for higher-volume projects.

Manufacturing

For more than 100 years, the manufacture of lenses has remained essentially unchanged. While these conventional methods utilize relatively low-cost machinery, they are also very labor intensive and require highly skilled craftsmen. With recent innovations in computer numerically controlled (CNC) machines, faster and less labor-intensive manufacturing methods are now viable options over conventional methods. From prototyping to high-volume production, automated grinding and polishing technologies are now available for lens fabrication.

Although these new technologies are more efficient and provide more reliable production, they require a significant initial capital investment. In addition, there are some situations where conventional methods are simpler to use and more cost effective. To understand the practical applications and benefits of each type of lens fabrication, brief descriptions of the manufacturing methods follow.

Conventional Lens Fabrication

Conventional lens fabrication (Fig. 18.3) begins with a plano-plano disk of glass or a near form-molded lens blank. The blank is placed into a chuck that rotates around the mechanical center of the glass disk. A ring tool with embedded diamonds removes bulk material and grinds down the top surface of the blank. This process gives the lens blank a spherical shape and a coarse surface finish. This surface has significant subsurface microfractures, which must be removed by loose abrasive lapping at a later stage in the manufacturing process. The lens blank can then be flipped and its second side ground to near net shape using the same process. This overall process is called *generating* because the end result is the generation of a blank in the shape of the final lens.

Fine Grinding and Polishing

To prepare for the fine-grinding and polishing process, the perimeter of the lens blank is wrapped with tape to create a reservoir. Molten pitch is

Figure 18.3
Conventional
Generation

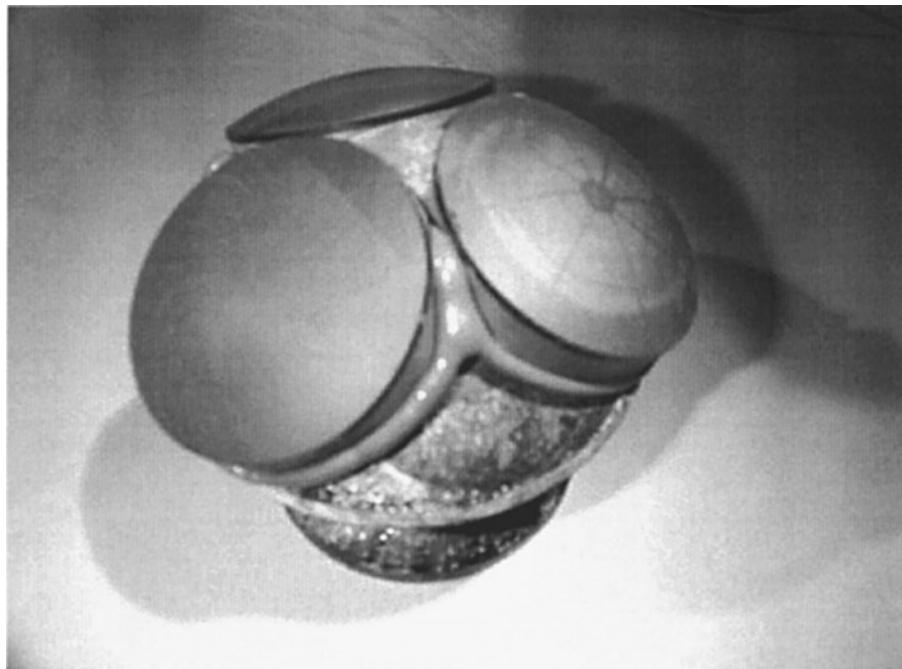


poured onto the surface of the lens, filling the reservoir. The pitch is allowed to cool at room temperature until a solidified pitch layer, called a *pitch button*, is developed.

The next step is to arrange the lenses in a circular pattern to be processed as a group called a *multiple block*. Multiple blocks are assembled by laying the buttoned lenses into a tool, which has a radius approximately equal to the design radius. Now, with the generated spherical surface down and the pitch button side facing up the array of lenses is ready to receive the blocking tool. This heated metal tool is placed in contact with the pitch buttons, allowed to melt into the pitch, and then quickly cooled to room temperature. The resultant “block of lenses” is then ready for loose abrasive lapping.

The purpose of loose abrasive lapping, often referred to as grinding, is to remove the residual subsurface damage that was incurred during the generating process. The block of lenses (Fig. 18.4) is fine ground, with loose abrasive grains mixed with water. Grinding is a step-down process that begins with large grains and continues with sequentially smaller and smaller grains. Grain sizes typically range from 30 to 5 μm . At this

Figure 18.4
A Block of Lenses



point in the manufacturing process, the operator is trying to achieve two goals: (1) a spherical surface very close to the design radius and (2) no subsurface damage. It is important for the optician to be aware of the abrasion resistance of the glass in order to control center thickness while minimizing subsurface damage. To achieve a thickness within the center thickness tolerance, a certain amount of material (on the order of tens of micrometers) is left on the lens for removal during polishing.

The lens is polished to the specified radius of curvature, spherical irregularity, and cosmetic finish by using a soft-pitch lap pressed to the desired radius and rotated about the spherical lens surface while a cerium oxide-polishing slurry is applied. The radius of the lens is controlled with a test glass or test plate of known radius. The lens is compared to the test plate by direct contact and/or evaluating the fringes of the Fizeau interferometric test. This test also gives the optician the ability to measure spherical irregularity, which is the maximum allowable perturbation of the spherical wavefront. The cosmetic requirements for the lens dictate the maximum allowable surface imperfections such as scratches, digs, and chips.

Conventional Centering

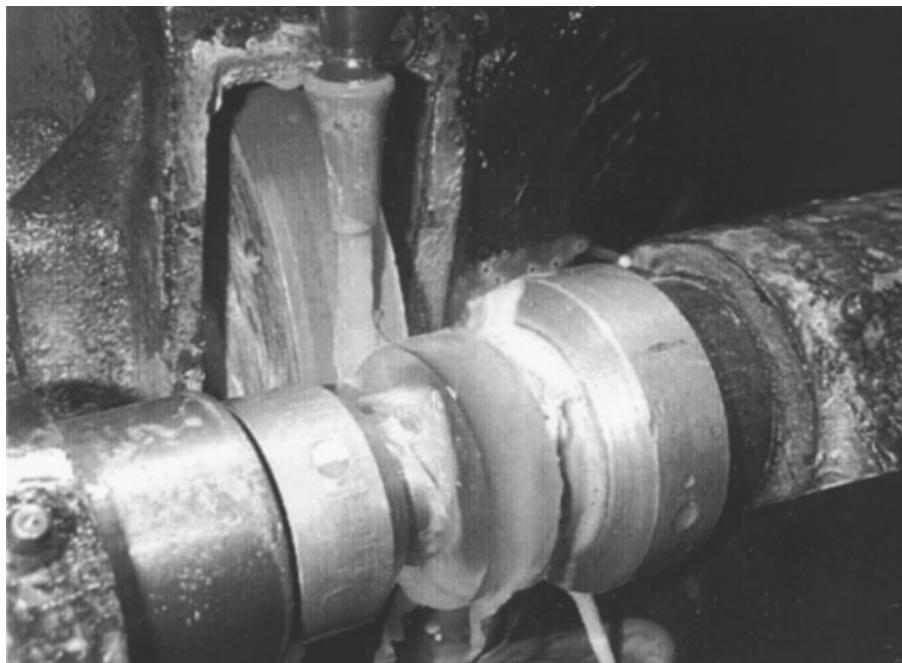
Once both sides of the lens are polished, the lens is centered by precision grinding the edge of the lens on a special lathe (Fig. 18.5). This process accomplishes two tasks. First, the lens is ground to its final diameter. Second, the optical and mechanical axes of the lens are made coaxial with one another. This is also the point at which any flats or special mounting bevels are ground onto the lens.

Once the lens is centered, manufacturing is complete. The lens is cleaned and inspected for quality. If it is satisfactory, the lens will be delivered either uncoated or with an antireflection coating. If the lens is not satisfactory, it is returned to one or more of the steps in the process to be corrected. If the lens cannot be reworked to meet the required specifications, it is scrapped.

CNC Lens Fabrication

Recent advancements spearheaded by the Center for Optics Manufacturing (COM) at the University of Rochester in Rochester, New York, have led to the development of equipment and processes that enable the

Figure 18.5
Conventional
Centering



optician to perform a variety of operations on computer-controlled machines—processes called *CNC lens fabrication*.

The equipment combines the accuracy of multiaxis CNC motion control with robust machine designs that are faster, more versatile, and more precise than conventional machines. This automated process minimizes part-handling and transfer errors, which are prone to happen with the more manual conventional process. It also enables the optician to generate precision surfaces that are precentered to final diameter and ready to polish. The precision spindles yield little subsurface damage, which reduces polishing time and shortens overall production time. An added benefit of using this equipment is that the lens can be shaped to precise complex dimensions during the generation sequence.

Once the lens has been precision generated, it is polished to meet all the requirements for surface accuracy and cosmetics. Polishing may be done utilizing the conventional process described earlier in the “Conventional Manufacturing” section of this chapter or with a new

Figure 18.6
Deterministic Grinding with CNC Machine

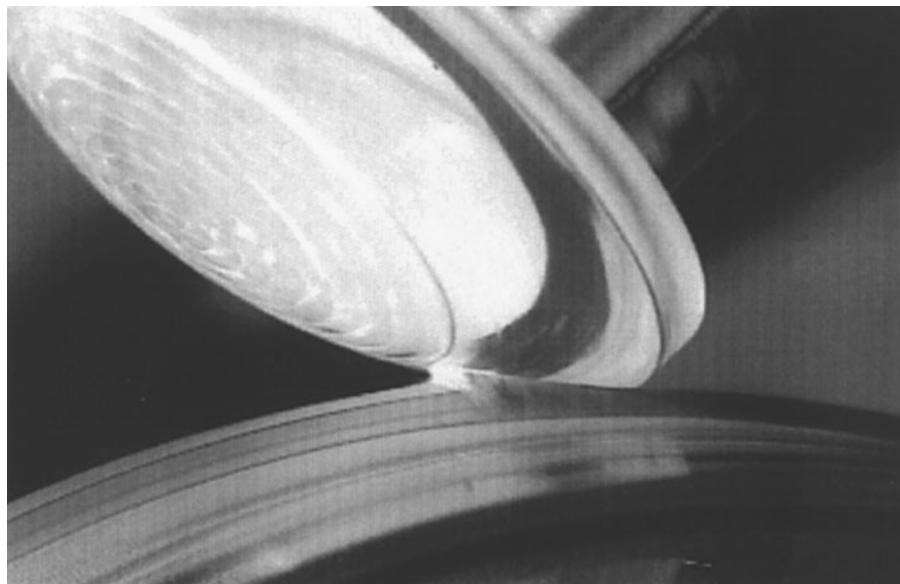


CNC polishing machine (Fig. 18.6). Determining which polishing method is most appropriate depends on geometry of the lens as well as the quantity being produced. For example, if the lens has a relatively long radius of curvature, then conventional polishing of a multiple block may be most cost effective.

It is important to note that this is a two-machine process with very good process control. For most applications the lens fabrication is complete. However, for high-precision applications COM has developed a complementary machining technology that can significantly improve the surface figure of the lens.

This new technology uses a unique fluid that is magnetically manipulated to deterministically remove material from the lens. The process is called magnetorheological finishing (MRF) (Fig. 18.7). The magnetorheological (MR) fluid stiffens as it passes through a magnetic field, thus forming a temporary finishing surface or polishing pad. The MR fluid carries polishing slurry that is presented to the lens surface in a precisely controlled pattern by varying the magnetic field's strength and direction. Since fresh abrasive is continuously delivered to the polishing zone, heat and debris are constantly removed. This

Figure 18.7
Magnetorheological
Finishing (MRF)



process reduces cycle times and is capable of producing fractional wave-surface irregularity.

The development of CNC machine technologies led directly to the capability to fabricate precision aspheric lenses in brittle materials, for example, optical glass. Robust CNC machines are able to profile grind complex rotationally symmetric shapes defined by polynomial equations (Fig. 18.8). This development effort continues today. Commercially viable processing methods are being developed for conformal optics. Conformal optics is loosely defined as nonrotationally symmetric, such as a saddle or a toroid. In fact, processing methods for conformal topics have progressed so far that testing the finished optic is often more challenging than making it.

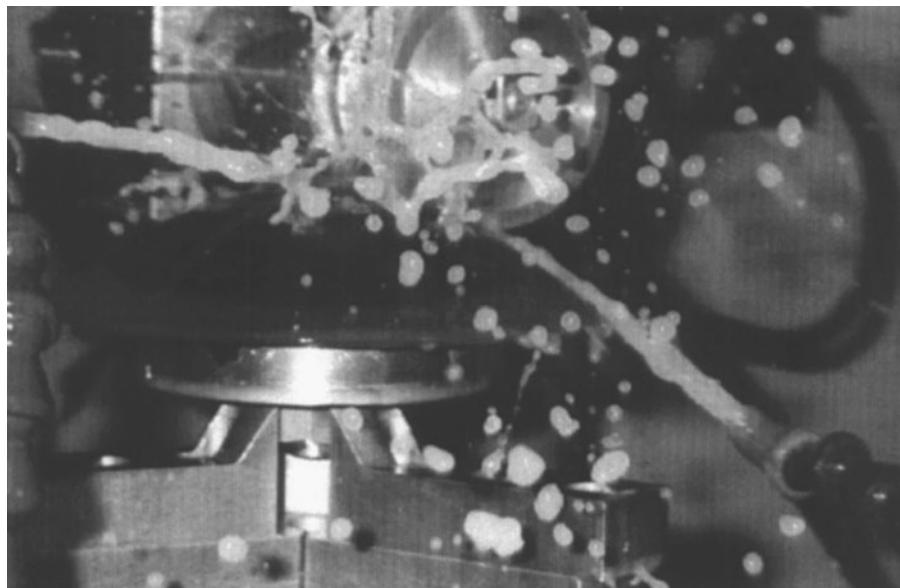
For more information regarding these new technologies please visit one or more of the following web sites:

www.Optipro.com

www.QED.com

www.optimaxsi.com

Figure 18.8
Precision Fabrication
of Aspheric Lenses via
CNC Machining



Special Fabrication Considerations

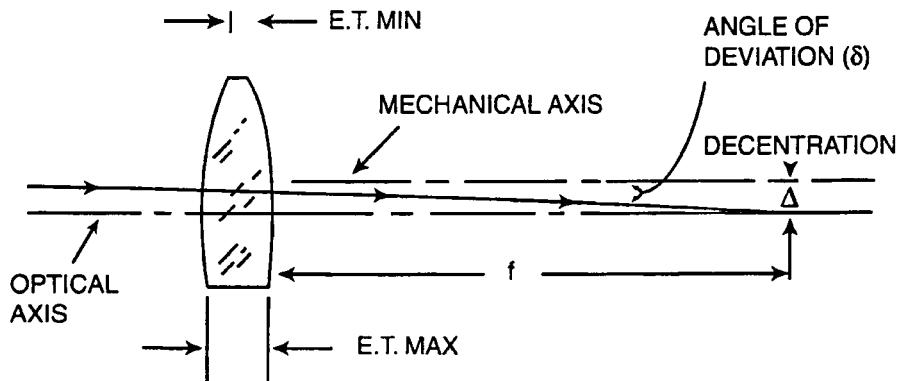
Centering Tolerance

Centering tolerance is a complex optomechanical parameter that is frequently misinterpreted. For example, 1-arc min edge thickness difference (ETD) may be reasonable for a 50-mm-diameter lens, but a 6-mm-diameter lens with this tolerance requires centering to 0.003 mm ETD which is extremely difficult. Figure 18.9 shows the relationship between the optical and mechanical axes and the decentration and angle of deviation in a decentered lens. Table 18.2 demonstrates the relationships between different wedge specifications. These equations provide conversions from one tolerance designation to another. The equations work well for most lenses, but lose accuracy with meniscus lenses as they approach concentricity.

If the tolerance analysis indicates that surfaces must be controlled to a few micrometers, then precision potting of the finished components should be considered. *Precision potting* refers to active alignment of the optic axis to the mechanical axis within the mounting cell. Since it is difficult to center lenses to ETDs of less than 10 μm , assembly techniques

Figure 18.9

Relationship Between Optical and Mechanical Axes and Decentration and Angle of Deviation in Decentered Lens



have been developed to provide submicrometer alignment. For most optical systems, it is not beneficial to put unusually tight constraints on the lens because the housing in which it will be mounted typically will have more error than the lens.

Clear Aperture

Clear aperture is a specification dimension. It should provide enough aperture for light rays to pass through; however, many problems can result from the clear aperture being specified too close to the outside diameter of the lens. For example, achieving fractional wavelength surface quality will be difficult due to edge roll-off in polishing. In addition, during coating, the lens is held mechanically in a fixture above the coating

TABLE 18.2 Centering Tolerance Specifications

	Deviation (Dv)	Edge Runout (ERO)	Surface Runout (ETD)
Deviation (Dv)		$Dv = 1720 \times ERO/f$	$Dv = 3440 \times (n - 1) \times ETD/D$
Edge runout (ERO)	$ERO = D \times Dv/3440 \times (n - 1)$		$ERO = 2 \times f(n - 1) \times ETD/Dv$
Surface runout (ETD)	$ETD = D \times Dv/3440 \times (n - 1)$	$ETD = D \times ERO/2 \times f(n - 1)$	

Note: Dv = deviation (in minutes); ERO = edge runout; ETD = edge thickness difference; D = diameter; f = focal length; n = material index of refraction (same value used to calculate focal length)

source. Therefore, it is important to have sufficient clearance between physical diameter of the optic and clear aperture. Ideally, a clear aperture-to-diameter difference should be at least 2.0 mm, or 5% of the aperture, whichever measurement is greater.

There is an alternative when the clearance is not adequate for the coating tooling: the lens may be coated before edging. This option is not desirable, however, because the coating will be at risk during the edging process.

Thickness Tolerance

Thickness tolerance is more difficult to achieve on softer glasses that are less resistant to abrasion. When tight thickness tolerance is required along with very stringent cosmetics and fractional wavelength irregularity, the optician must allow the right amount of excess material to accommodate for grinding and polishing of the lens. This causes a wider range of center thickness, which may produce lower production yields. As a result, it may be necessary to start more pieces to account for the expected losses.

Sag Tolerance

Sag tolerance is sometimes specified as the desired clear aperture. This is a difficult feature to measure. A better method is to compute the sag as a function of the clear aperture and the radius. This yields an *axial height*—the on-axis distance from the plane of the sag face to the spherical surface—which can be easily and accurately measured. If the sag face is used as a mounting surface, then the tolerances for the sag and center thickness are cumulative. If the sag is not a mounting surface, then it should be identified as a reference (Ref) surface in order to reduce cost.

Radius Tolerance

Radius tolerance is used to specify the allowable radius measurement deviation from nominal for the test plate (that is, spherical reference tool) that will be used for lens production. For precision optics this measurement is typically 0.1% of the nominal radius and not less than 10 μm for short radii. The optical designer should be aware there are no industry standards for radius measurement and that absolute radius measurement

is not possible. If five different optics manufacturers measure a test plate, then there will be five different readings. For radii less than 1000 mm, the variation will be on the order of a few micrometers. For radii over 1000 mm the variation could be several millimeters. Researchers at The National Institute for Standards and Technology are working toward a solution to this problem.

Power Tolerance

Power tolerance is a measure of the deviation from the chosen test plate. This ensures consistency among a group of lenses. In other words, each lens will match the test plate within the power tolerance. From the designer's perspective, the radius tolerance and the power tolerance are cumulative. The original purpose of the power tolerance was to indicate the maximum number of power fringes for which the irregularity fringes could be counted. For example, in order to see two fringes of irregularity, the maximum number of power fringes is 10 fringes; for one fringe irregularity the maximum is five fringes. However, automated interferometric metrology is reducing the need to rely on traditional test plates.

Surface Irregularity

Surface irregularity is a measure of the deviation from a perfect sphere. It is not only a function of the operator's skill and expertise but also a function of the process geometry. As a general rule, multiple blocks with more pieces will have less irregularity than three spots or singles. The irregularity of lenses processed on multiple blocks will have a tendency to be cylindrical in nature while lenses processed as singles will have a symmetric aspheric profile shape, usually like a sombrero. There are certainly exceptions to this rule, but the general shape of the irregularity will follow these tendencies. Irregularity is defined very well by ISO 10110. See Fig. 18.10 for more information.

Aspheric Lens

Aspheric lens manufacturing technology has progressed rapidly over the past few years. The pace of this progress is limited somewhat by the difficulty in measuring aspheric profiles that include up to sixteenth-order

Figure 18.10

Excerpt from ISO 10110-5 Optics and Optical Instruments. Preparation of Drawings for Optical Elements and Systems—Part 5: Surface Form Tolerances

3.5.2 Irregularity

The *irregularity* of a nominally spherical surface is a measure of its departure from sphericity.

The value of the irregularity of an optical surface is equal to the peak-to-valley difference between the optical surface under test and the approximating spherical surface.

3.5.3 Rotationally Symmetric Irregularity

Surfaces which are rotationally symmetric, but do not have the desired shape, are said to have *rotationally symmetric irregularity*. This error is the rotationally symmetric part of the irregularity function (see subclause 3.5.2).

In order to determine the value of the rotationally symmetric irregularity, d is first necessary to determine the rotationally symmetric aspheric surface which best approximates the surface under test.

terms. Using bonded diamond-tool generation for brittle materials (for example, glass), convex aspheres are usually easier to fabricate than concave surfaces. The computer-controlled machines can process complex shapes irrespective of best-fit sphere. In contrast, single-point diamond-turning machines can produce convex and concave surfaces on plastic and crystalline materials. However, small departures from best-fit sphere are preferred. The manufacturing cost for aspheres is typically 2 to 5 times that of spherical lenses with short radii. The generally accepted method for metrology is surface profiling to an accuracy of $\pm 0.1 \mu\text{m}$. Aspheric form error on the order of $50 \mu\text{m}$ may be good enough for a condenser lens, while a precision quality focusing lens would require $\pm 1 \mu\text{m}$ tolerance. Greater precision is possible with interferometric testing, which often requires the fabrication of a special null lens.

Bevels, Chamfers, and Break Edges

Bevels, chamfers, and break edges are machining features utilized at the corners of a lens to help prevent edge chipping. Bevels should be specified whenever the included angle of two surfaces on an optic is less than 155° .

Cosmetic Tolerances

Cosmetic tolerances are well defined in MIL-O-13830 and ISO 10110. Most cosmetic inspection of lenses is still done visually by comparing the

lenses to scratch-dig reference pieces. Alternatively, defects can be evaluated and categorized using a measuring microscope.

Antireflection Coatings

Antireflection coatings are a significant cost driver and can be reduced with minimal design effort. The most economical solution is to coat all surfaces with a single-layer MgF_2 coating. This enables the lenses to be coated all in one run (depending on size and quantity). Single-layer MgF_2 coating will yield about 1.5% reflection for each low-index surface and less than 1.0% reflection for each high-index surface. For multielement systems, specifying different coatings within the system can minimize coating costs. For example, the high-index glasses may be coated with MgF_2 , while a broad band antireflection (BBAR) coating is applied to the low-index material. As a result, coating cost may be reduced because only the low-index glasses are receiving multilayer BBAR coatings, which are more expensive than MgF_2 coating. When using this approach, the designer should consider that BBAR coatings are index dependent. The coater will batch lenses by index—less than 1.60, 1.60 to 1.70, and greater than 1.70. Using glasses within two of these ranges instead of all three will reduce coating costs.

Blocking Quantities

Blocking quantities are a function of the relationship between the radius and the diameter of a lens. The graph in Fig. 18.11 reveals the relationship between radius, diameter, and blocking quantity. For example, lenses with a radius-to-diameter ratio of less than 0.84 will process as a single, a ratio of 0.84 to 1.04 will run three pieces to a block, and so on. The more surfaces per block, the lower the cost per lens to process it. The diameter of the parts and the capabilities of the manufacturer put additional parameters around the number of pieces that can be produced at one time.

Concentric Lenses

Concentric lenses (Fig. 18.12) create a problem with centering accuracy. Since the centers of curvature for both surfaces are close to one another,

Figure 18.11
Blocking Graph

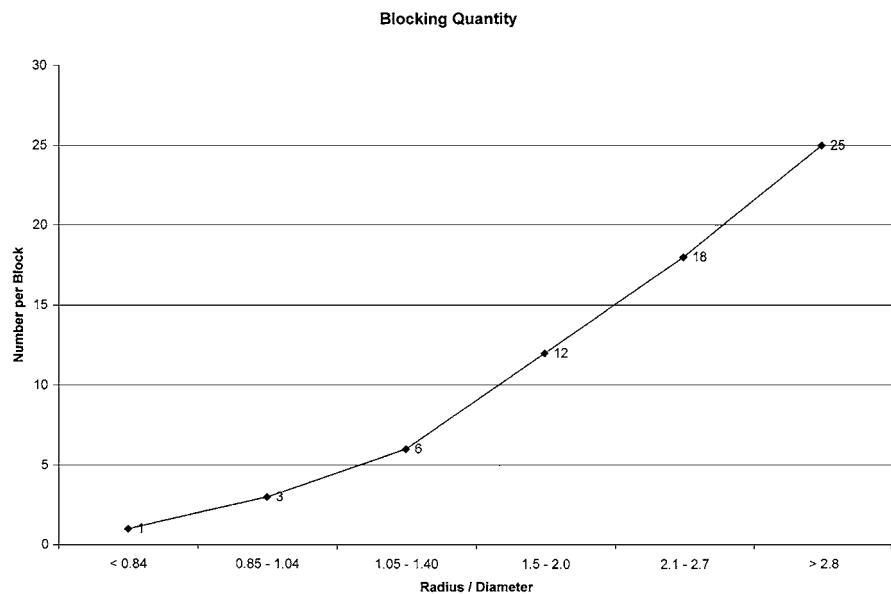
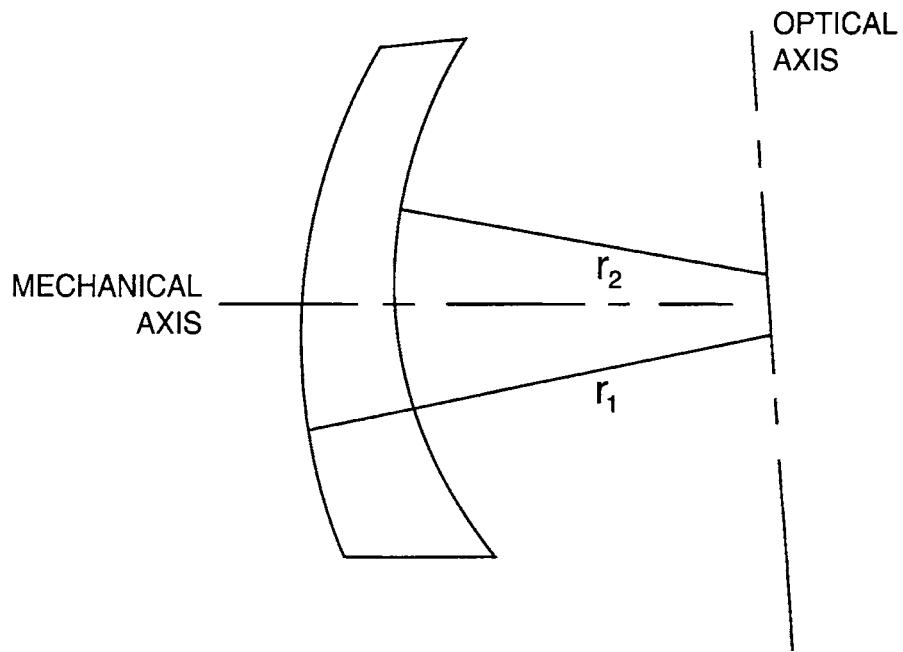


Figure 18.12
Concentric Lens



THE TWO AXES CANNOT BE MADE COAXIAL

the optician is not able to remove much residual wedge in the centering process. When the concentric lenses have weak curves and the lenses are processed as a multiple block, special care must be taken during blocking and grinding to prevent wedge in the part. In general, it is best to process concentric lenses individually on CNC equipment, where the tolerances can be well controlled.

Hemispheres and Hyperhemispheres

Hemispheres and hyperhemispheres are difficult to process because the polishing tool must rotate beyond the waist of the lens. This requires specialized machines and tooling. Small convex hemispheres are often made by modifying spheres to the desired shape. The use of concave hemispheres should be avoided whenever possible due to manufacturing difficulties associated with these shapes. It is important to note that designing and applying an antireflective coating for all angles of incidence presents another set of challenges such as special coating textures to apply uniform antireflection coatings.

Aspect Ratio

Aspect ratio is the relationship of center thickness to diameter. The higher the ratio, the higher the probability that the glass will distort during processing. The distortion is a function of thermal stress caused by the application of heated pitch and its subsequent cooling. After polishing, the lens is deblocked from the pitch and the stress is relieved. Lenses with extremely thin centers or thin edges are prone to develop surface irregularities during processing. Ideal aspect ratios are less than 6:1 for precision optics with one half-fringe irregularity. Aspect ratios greater than 10:1 will be more problematic and therefore more costly. There is also a greater likelihood for surface deformation from mounting and the assembly process.

Thin Edges

Thin edges can occur when there is a strong convex surface on at least one side of the lens. When the edge is thin (<1 mm), it is more fragile

and prone to chipping, and the optician is able to protect only the edge with a minimal bevel. Thin edges cause flaking out of glass particles during polishing, which leads to difficulty in achieving good cosmetic surfaces. A thin-edge lens is also difficult to hold in place for testing on an interferometer because the slightest amount of pressure causes the lens to distort.

Segmenting

Segmenting refers to special mechanical shaping. It is difficult to polish lenses of noncircular geometries. Therefore, if segmenting is required, the manufacturer will usually perform this step last. Unfortunately, all of the value (material and labor) has already been invested in the lens, which inherently makes this a high-risk process.

Edge Blackening

Edge blackening of the lens helps reduce scattered light and often improves contrast and signal-to-noise ratio. Permanent black ink that is water and alcohol resistant is easy to apply and does not cause mechanical buildup on the surface. Lacquers and epoxies are more opaque; however, they are more difficult to apply and add tens of micrometers to the diameter of the lens. Epoxy is the most durable option, and if factored in during the design of the lens, it will not negatively impact the finished diameter of the lens.

Component Testing

An important consideration before manufacturing begins is *component testing*, which verifies that all parameters of the lens can be measured to the desired accuracy. Inspection data should be provided with all prototype components. In the event that an optical system does not perform as predicted by its design, the system can be computer modeled using the actual test data. In production, it may be helpful to perform inspection in compliance with the military specification MIL-PRF-13830B for a prespecified acceptable quality level (AQL).

Cemented Doublets

Cemented doublets can enhance optical system performance without decreasing light throughput. There are many methods for making doublets and the optimal choice will require some design consideration, including thickness tolerance, surface irregularity, and assembly.

Doublets yield some thickness flexibility for the designer. Most doublets are made from a flint and crown lens and have optical adhesive indices of about 1.5, similar to the crown glass. When tight thickness control is needed on a doublet, rather than give half the tolerance to each half of the doublet, the designer may be able to give the whole tolerance to each half and then have the optician match the thickness of each half before cementing them to make the doublet fall within the tolerance band. This can be a cost-effective solution to controlling doublet thickness.

All optical adhesives have some amount of shrinkage due to curing. This shrinkage can cause deformation of the lens elements and compromise the irregularity of the polished surfaces. Avoiding thin lenses in doublets and selecting a low-shrinkage adhesive help minimize this effect.

The assembly method for a doublet will depend on the wedge tolerance (see Table 18.3). The simplest approach is to center each half of the doublet to the same diameter and use the edges of the lens for alignment. This method is best suited for lenses greater than 15 mm in diameter. Another method is to center one half, the base lens, to a precision diameter and center the other half, the floater, to a smaller and less precise diameter. Then,

TABLE 18.3

Centered Doublet Guidelines

Alignment Method	Mechanical Consideration	Precision (arc min)
V-block aligns the diameters of the two lenses to be cocylindrical.	Precision center the lenses to the same diameter and desired wedge tolerance.	6
Bell clamping aligns the polished surfaces to be coaxial by mechanical positioning.	Precision center the base lens and center the floater to a smaller diameter.	3
Active alignment aligns the polished surfaces to be coaxial by visual interactive positioning.	Precision center the base lens and center the floater to a smaller diameter.	<1

referencing on the base lens, the optic axis of the floater can be aligned. In special cases the doublet can be built and centered as the final process. This is a very high-risk process and should be avoided when possible.

Relative Manufacturing Cost

In addition to the design considerations already described in this chapter, there are several other variables that can significantly impact the relative manufacturing cost of lenses. For example, the tolerances given to manufacturing specifications can lead to additional costs being incurred during manufacture of the lens. Other variables that may influence cost are the aspect ratio and the preferred delivery time.

In the mid-1970s, J. Plummer and W. Lagger wrote an article for *Photonics Spectra* entitled “Cost Effective Design.” The article contrasted the effect of manufacturing tolerance on the cost to make a lens. The chart from that article, represented in Table 18.4, has been updated to detail the cost impact of several variables for a manufacturing process that utilizes the newer deterministic microgrinding technology.

These relative costs are not cumulative, but are clearly interrelated as previous comments in this chapter’s discussion have indicated. The total cost impact of several factors would be a complex mathematical function, and would vary from shop to shop, depending on the capabilities and strengths of each shop.

Sourcing Considerations

Every project has specific goals, such as to bring a new product to market before the competition, to develop a new capability, or to reduce manufacturing cost. In order to be successful, the project manager must determine the priorities among price, quality, and timeliness. The manager must then communicate those priorities to everyone involved with completing the project. The following guidelines are offered for consideration in achieving cost, quality, and delivery goals:

- To minimize cost and delivery time, buy from a catalog whenever possible. At the same time, keep in mind that custom lenses are often required in order to achieve a desired optical performance.

TABLE 18.4

1999 Relative Manufacturing Costs Using Deterministic CNC Processing

	Variable	→ More Difficult → →				
		± 0.10	± 0.05	± 0.025	± 0.0125	> 0.0075
Diameter (mm)	\$100	\$100	\$102	\$105	\$125	
	\$100	\$103	\$115	\$140	\$200	
Thickness (mm)	± 0.20	± 0.10	± 0.05	± 0.025	$> \pm 0.0125$	
	\$100	\$103	\$115	\$140	\$200	
Stain	<2	2	3	4	5	
	\$100	\$103	\$110	\$140	\$175	
Cosmetics (Scr-Dig)	80-50	60-40	40-20	20-10	10-5	
	\$100	\$100	\$120	\$150	\$250	
Test (fringes)	5-2	3-1	$2\frac{1}{2}$	$1\frac{1}{4}$	$1\frac{1}{2}-1\frac{1}{8}$	
	\$100	\$105	\$125	\$175	\$250	
Wedge (arc min)	3	2	1	$\frac{1}{2}$	$\frac{1}{4}$	
	\$100	\$105	\$110	\$125	\$150	
Doublets (arc min)	6	3	2	1	$< \frac{1}{2}$	
	\$100	\$105	\$110	\$150	\$200	
Aspect ratio	<10:1	15:1	20:1	30:1	50:1	
	\$100	\$120	\$175	\$250	\$350	
Delivery time (weeks)	8	6	4	2	1	
	\$100	\$110	\$130	\$170	\$200	

- To minimize risk on a project (that is, maximize the potential for good quality and on-time delivery), use domestic manufacturers for prototyping and preproduction. Optics manufacturers in the United States have superior manufacturing capabilities for rapid prototyping, high-precision optics, computer-generated holographic (CGH) and diffractive optical elements (DOEs), laser optics, precision glass aspheres, polarizers, complex optical coatings, and much more.
- Rapid prototyping can significantly minimize cost and delivery time. Some projects are very time sensitive and optical components become the pacing item. Typical delivery time for rapid prototyping is 8 to 10 weeks. Seek a manufacturer with a proven

track record. Several manufacturers have developed the ability to expedite the manufacturing process to achieve shipment of coated optics within a few days. This service may require a premium on the standard price.

- To reduce cost with relatively low risk, seek a domestic importer with an established offshore facility that has the ability to test and certify product quality. Or for the lowest price, consider working directly with an offshore supplier. However, this is quite risky if you don't have the appropriate metrology to verify the product quality.

Conclusion

This chapter presents a great deal of information to help the designer select attributes and tolerances based on manufacturing considerations. Perhaps the most useful summary is a reference chart that provides a list of reasonable or typical manufacturing tolerances for commercial quality and precision quality lenses (see Table 18.5). This chart is intended as a

TABLE 18.5

Typical Manufacturing Tolerances

OPTIMAX	Commercial Quality	Precision Quality	Manufacturing Limits
Glass quality (n_d)	± 0.001	± 0.0005	Melt controlled
Diameter (mm)	$+0.00/-0.10$	$+0.000/-0.025$	$+0.000/-0.010$
Center thickness (mm)	± 0.150	± 0.050	± 0.010
Sag (mm)	± 0.050	± 0.025	± 0.010
Radius (%)	± 0.2	± 0.1	± 0.025
Power-irregularity (fringe)	5–2	3–0.5	1–0.1
Aspheric profile (μm)	± 25	± 1	± 0.5
Wedge lens (TIR, mm)	0.050	0.010	0.005
Prism angles (TIA, arc min)	± 3	± 0.5	± 0.1
Bevels (maximum face width at 45° , mm)	1.0	0.5	No bevel
SCR-DIG	80–50	60–40	10–5
AR coating (average R)	$\text{MgF}_2, R < 1.5\%$	BBAR, $R < 0.5\%$	Custom design

guideline and assumes a 50-mm-diameter BK-7 lens. The manufacturing limits are not absolute, but represent a pain and/or cost threshold. Job-specific tolerances may vary depending on component size, shape, glass material, and preferred delivery time.

Once a lens has been designed and toleranced, manufacturing drawings are utilized to convey the lens requirements to the optician. Examples of a conventional manufacturing print and a drawing that complies with ISO standards follow (Figs. 18.13a and b). For more information, see Part 10, "Table Representing Data of a Lens Element," within ISO 10110 for element drawings ("Optics and Optical Instruments: Preparation of Drawings for Optical Elements and Systems").

	Radius	Radius tolerance	Power/ Irreg.	Clear aper. diameter.	Edge diameter	Diameter tolerance	CT	Thickness tolerance	Wedge
S1	13.589 CC	TPF	5/3	20.6	31.18	± 0.05	4.45	± 0.05	0.05 TIR
S2	64.643 CX	TPF	5/3	29.0					

Notes:

1. All dimensions in inches.
2. Material: Optical glass per MIL-G-174 type BK7
Schott N.o. 517642
 $n_d = 1.5168 \pm 0.0005$ $V = 64.2 \pm 0.8\%$ striae grade
A, fine anneal
3. Surfaces marked "P" polish to power/irregularity indicated
4. Manufacturer per mil-0-13830
5. Surface quality 60 - 40
6. Surfaces marked "C" coat with high efficiency coating with average reflectance per surface $\leq 0.5\%$ from 420 - 680 nm
7. Surface marked "G" fine grind and blacken with no buildup
8. Bevel edges at 45 degrees to 0.5 max face width
9. Diameter to flat is 19.952 (REF) with surface sag of 5.977 ± 0.05 on surface S1

The diagram shows a lens element with two curved surfaces, S1 and S2. Surface S1 is labeled 'P, C' and surface S2 is labeled 'S2 P, C'. An arrow points to the edge of the lens with the label 'G'.

DR	
CHK	
APPD	
SCALE - 2.5:1	

Sample Element Print

Figure 18.13a
Conventional Lens Manufacturing Print

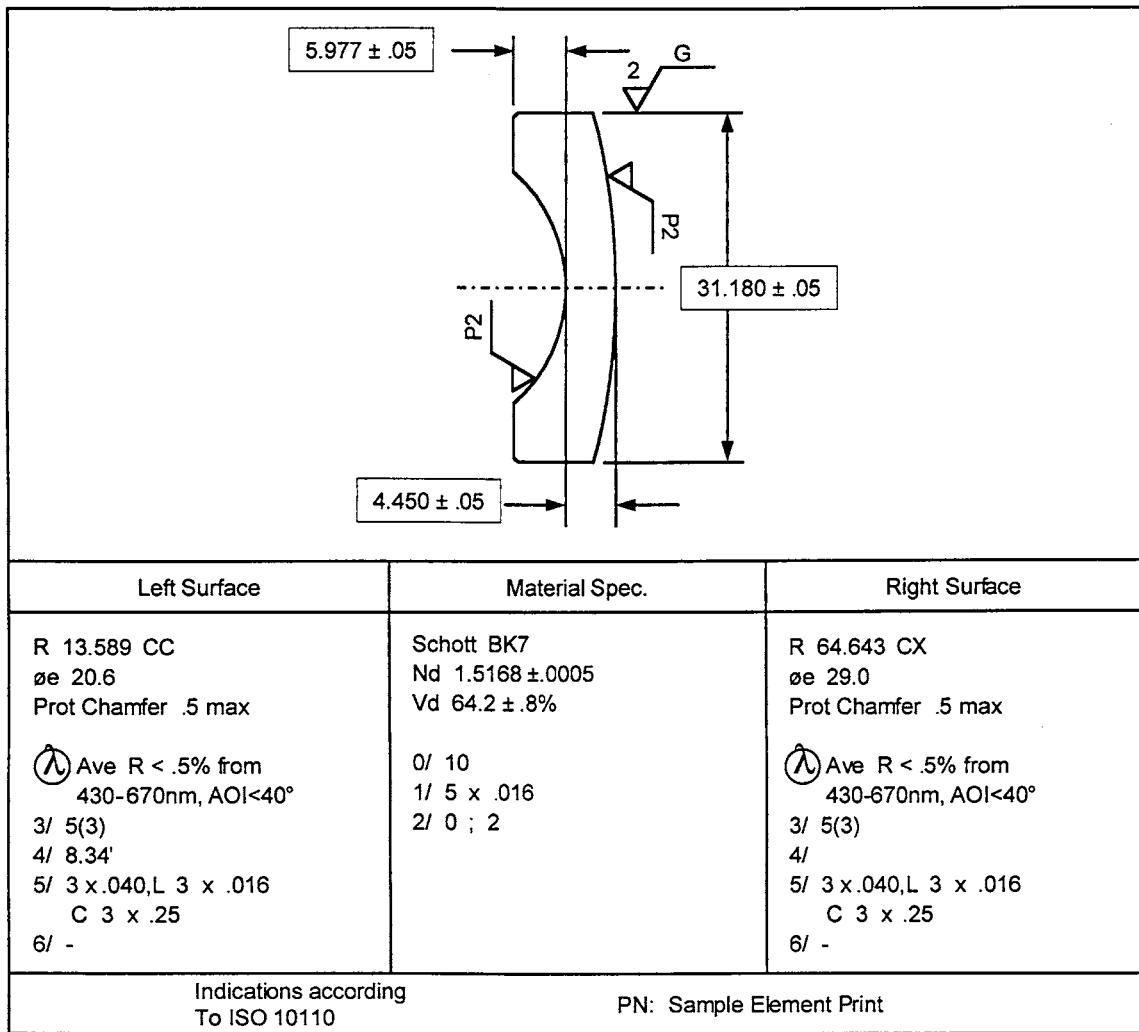


Figure 18.13b
ISO Lens Manufacturing Print

Polarization Issues in Optical Design

Introduction

In geometric optics light is considered to propagate as bundles of energy moving in straight line paths called *rays* which bend at interfaces between media with differing refractive index. Where necessary the notion of the wave nature of light is included, which allows for the interference of light beams, whose amplitudes add when they are superimposed. An additional level of sophistication is that of polarization: The ray of light possesses an electric field vector that lies in the plane normal to the direction of wave propagation. To the extent that this vector direction is predictable the light is called polarized, and to the extent that this vector direction is unpredictable the light is called unpolarized.

The study of polarized light and polarization dependent optical effects, not surprisingly, involves a fair amount of optical physics and electromagnetic theory. The physical principles of most optical polarization phenomena have been understood for many decades, but applying the physics of polarization to optical systems can be tricky. However, most polarization issues that arise in optical design can be adequately addressed with simple models, without the need to resort to vector diffraction theory.

The first polarization question facing the optical engineer is whether in a given application he or she should even care about polarization. A polarization critical optical system is one in which the behavior of polarized light is critical to the successful performance of the system. In a liquid crystal projection display, for example, the polarization behavior

of the liquid crystal material, polarizers, and surrounding lenses, mirrors, and filters is clearly crucial to attaining a high-contrast image. This highly polarizing optical system is polarization critical. However, even in systems containing no highly polarizing elements, the weakly polarizing nature of many optical elements can in some cases result in unacceptable performance.

The purpose of this chapter is to introduce concepts of optical polarization to the optical engineer and designer with an emphasis on issues that are most likely to arise in a wide range of optical applications. It should help the engineer to identify polarization critical optical systems and to analyze their performance. It cannot take the place of specialized texts on polarization topics, which give a far more complete discussion of electromagnetic theory, polarization measurement in theory and in practice, polarization aberration theory, and polarization ray tracing.

The first section of the chapter introduces polarized light qualitatively in order to present the principles of polarization and polarization phenomena without mathematical baggage. Those needing to delve into quantitative polarization analysis will require some understanding of two mathematical tools—the Jones calculus and the Mueller calculus—introduced in the second section. The third section provides brief descriptions of a range of polarization phenomena and devices. The fourth section discusses how to address polarization issues in the analysis of an optical system. The fifth section gives tips on how to minimize polarization problems in optical design. The sixth section discusses ways that polarization can be a help to the engineer in optical systems. Throughout the chapter we highlight **Polarization Pitfalls**, polarization issues that frequently cause confusion.

Introduction to Polarization

Polarized Light

Light is a propagating electromagnetic wave. It is a transverse wave, because the electric and magnetic field vectors lie perpendicular, or transverse, to the direction of the wave propagation. A sound wave, by comparison, is longitudinal because the pressure variation occurs in a direction along the direction of wave propagation. Within that transverse

plane the electric field vector may point in any direction. For a ray of light propagating horizontally, it is possible that the electric field vector will vibrate in the vertical direction, in the horizontal direction perpendicular to the wave vector, or in some mixture of the two.

If the electric field vector is constrained to vibrate in only the vertical direction, we say that the light is vertically polarized. If it vibrates only in the horizontal direction, we say it is horizontally polarized. There is nothing unique about vertical and horizontal: These are two special cases of linearly polarized light, whose electric field always falls along a fixed line.

It is not always the case, however, that light is linearly polarized. The electric field vector may also trace out a circle, or more generally, an ellipse in the plane transverse to the wave direction. In these cases the light is said to be circularly or elliptically polarized. Since motion in a circle can be either clockwise or counterclockwise, we also say that the polarization has a handedness, right or left.

Elliptical polarization is the most general fully polarized state. An elliptical polarization state can be described by its eccentricity (and handedness) and by the orientation of the major axis of the ellipse. If the eccentricity is near one, then the ellipse traces a shape that is very nearly a line, and we say that the light has a high degree of linear polarization. If the eccentricity is near zero, then the ellipse traces a shape that is very nearly a circle, and we say that the light has a high degree of circular polarization.

Combining two linearly polarized beams may result in another linearly polarized beam. Equal proportions of horizontal and vertically polarized light added together in phase result in light linearly polarized at 45° to the horizontal and vertical axes. However, combining two linear polarized beams may instead result in a circularly polarized beam. Equal proportions of horizontal and vertically polarized light added together 90° out of phase result in circular polarization: When the horizontal component is at a maximum, the vertical component is at a minimum. As time progresses the horizontal component decreases as the vertical component increases, rotating the vector formed by the superposition of the two components. In general, the result of combining two polarized beams depends not only on the magnitude and polarization of the constituent beams, but also on the phase difference between them.

Most natural sources of light, such as the spectral continuum of sunlight of magnetic regions, emit rays whose electric field vectors have no predictable orientation to them. Such light is called *unpolarized*. Of course,

at any instant in time the electric field has a direction, but the electric field vector of unpolarized light varies in an unpredictable fashion, so that it does not trace out a line or an ellipse over an observable period of time. The degree of polarization associated with a beam of light is zero if its electric field vector direction is unpredictable, and it is one if the electric field vector traces a fixed (predictable) ellipse. Light that tends to trace an ellipse, but not with perfect predictability, is partially polarized, having a degree of polarization greater than zero but less than one.

Polarization Behavior and Polarization Components

The human eye is nearly insensitive to the polarization state of light, so the eye is a poor polarization analyzer. Otherwise we would easily detect the pattern of linear polarization variation of a clear blue sky. Optical scattering is one of many phenomena which are polarization dependent. Ninety degrees from the sun the blue skylight is highly linearly polarized.

Polarized light may also be generated from unpolarized light by introducing a polarizer. The most common polarizer is a dichroic sheet polarizer such as those formerly made by Polaroid Corporation, which absorbs one component of linearly polarized light and transmits the other. The long chain polymers of sheet polarizers are doped with a conducting material, preferentially absorbing light of one polarization orientation. Good polarizers may also be made with narrowly spaced (subwavelength) metallic wires; these are called *wire grid polarizers*.

The reflection of light from a smooth surface is also polarization dependent. Sunlight that reflects from a road to a driver's eye tends to be more horizontally polarized than vertically polarized. Polarized sunglasses, which contain a sheet polarizer that passes only vertically polarized light, will block much of this glare. The same phenomenon allows you to easily determine the transmission axis of a sheet polarizer: Look through the polarizer at a beam of sunlight reflected from a road or an overhead light reflected from a horizontal table or a floor. Rotate the polarizer until the reflection is minimized. In this orientation the polarizer is blocking the reflection and transmitting vertically polarized light.

Some crystals produce different effects on incident light depending on the polarization state orientation. In fact, the phenomenon of polarization

was discovered when it was noticed that looking through a crystal of calcite one could see two images rather than one. One linear polarization state refracts with a different index than the orthogonal linear polarization state. The presence of two refractive indices (birefringence) makes crystals very useful for generating, analyzing, and converting polarization states. The orientation of the crystal axes determines the form of its polarization behavior.

Conversion of one polarization to another can occur through preferential absorption of one polarization state. It can also occur by changing the phase difference between components of the polarization vector. Elements that introduce polarization dependent phase delays are called *retarders*. The geometry of a birefringent crystal plate may be arranged such that the refractive index seen by vertically polarized light is different than that of horizontally polarized light. This occurs, for example, when the optic axis of a uniaxial crystal lies vertically in the plane of the polished faces of the plate. Therefore, the amount of time that it takes to pass through the crystal will be different for vertically and horizontally polarized light, and the phase change accumulated will also be different. If the accumulated phase difference is one-fourth of a wavelength of light, then the plate is called a *quarter wave retarder*. Quarter wave retarders have the interesting property that they can convert linearly to circularly polarized light and vice versa. Consider a ray polarized at 45° to the horizontal and vertical axes incident onto a quarter wave retarder. Before entering the retarder the oscillations of the horizontal and vertical components of the wave are perfectly in phase with each other. After entering the retarder, one component undergoes more oscillations than the other before exiting. This slower traveling component underwent one quarter additional phase oscillation at the exit of the crystal than its faster orthogonal partner. Then the two components of equal amplitude exit the crystal 90° out of phase, which is the identifying feature of circularly polarized light. Therefore, a quarter wave retarder converts incident light that is polarized at 45° to the crystal's slow and fast axes into circularly polarized light. Similarly, circularly polarized light incident onto a quarter wave retarder is converted into linearly polarized light whose orientation lies halfway between the slow and fast axis orientations.

POLARIZATION PITFALL A quarter wave retarder does not necessarily convert linearly polarized light into circularly polarized light. If a ray of light is polarized parallel to the fast (or slow) axis of the crystal,

then it propagates through the crystal with the fast (or slow) refractive index. The field experiences only one refractive index, and the propagation is no different from an isotropic piece of glass: The polarization state is unchanged. A quarter wave retarder converts linearly polarized light into circularly polarized light *only if* the incident light is linearly polarized at 45° to the fast and slow axes of the retarder. It will then be equally split into vector components along the slow and fast axis orientations of the crystal.

Similarly, a half-wave retarder induces one half of a wave of phase difference between the fast and slow polarization components of an incident beam. If an incident beam is linearly polarized at 45° to the fast and slow axes, then they reach a maximum at the same instant and a minimum at a later instant. After traversing the half-wave retarder, the 180° phase difference that has accumulated implies that when one component reaches a maximum positive value then the other component has reached a maximum negative value. The resulting combination of fast and slow components is still linearly polarized, but now at 135° to the fast and slow axes. The half-wave retarder essentially reflects the incident linear polarization state about its fast (or slow) axis. A half-wave retarder also reverses the handedness of a circularly polarized beam.

Retarders can modify the polarization state of incident light, but they do not affect its power or its degree of polarization. Diattenuators such as dichroic sheet polarizers can modify the polarization state of incident light, but they do so by absorbing, deflecting, or scattering power out of the beam.

Although we do not sense polarization directly, polarization optics is nevertheless important in science and technology. By measuring the polarization behavior of materials and objects we can learn more about them. Furthermore, by controlling polarization we can create useful devices. For example, at the heart of a liquid crystal television lies a pair of linear polarizers surrounding thousands of electrically addressed variable retarder pixels. The combined polarization behavior of the polarizers and retarders can result in the partial transmission of light through a pixel or in its complete blockage. The image built-up on an liquid crystal display (LCD) TV is achieved through exquisite polarization control. We discuss some details of the polarization control in these displays as well as other applications of polarization control in the "Polarization Analysis of an Optical System" section.

The Mathematical Description of Polarized Light

While mathematics is not required in order to obtain a useful understanding of polarization phenomena, the design and analysis of polarizing optical systems and of polarimeters demand a quantitative approach. This section provides a bare-bones introduction to the mathematics of polarized light. For those wishing to avoid the mathematical treatment, the subsequent sections of the chapter will still be qualitatively accessible. For those desiring a deeper and more complete mathematical treatment, excellent lengthier references are available.

The electric field of light is constrained to a plane perpendicular to the wave propagation direction. The state of polarization of a ray of light refers to the electric field orientation within that plane. To the extent that this direction is predictable, the light is said to be polarized, and to the extent to which this direction is random, the light is said to be unpolarized.

For example, consider the electric field vector associated with a ray of light of angular frequency ω propagating in the positive z direction:

$$\vec{E} = \hat{x}E_x \cos(kz - \omega t + \phi_x) + \hat{y}E_y \cos(kz - \omega t + \phi_y). \quad (19.1)$$

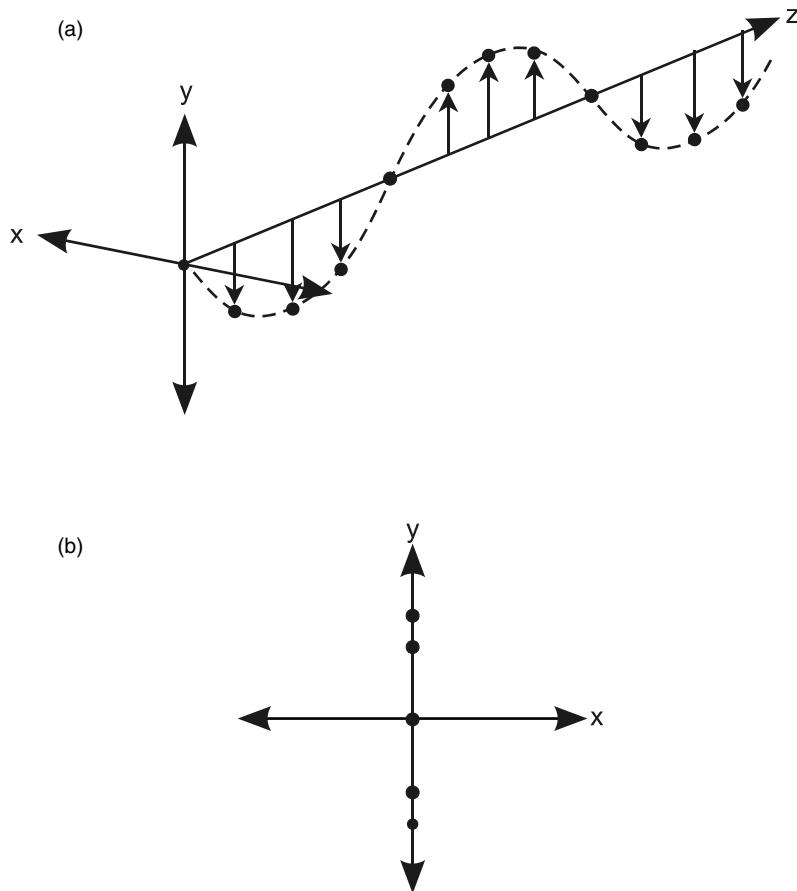
This ray is fully polarized, since the electric field orientation is fully determined. If $\phi_x = \phi_y$, or either E_x or E_y are 0, then the electric field vector at a particular value of z traces a line in the xy plane as the field oscillates. In this case the light is said to be linearly polarized. Figure 19.1 depicts light linearly polarized in the y direction.

If $E_x = E_y$ and $\phi_x - \phi_y = 90^\circ$ then the electric field vector at a particular value of z traces a circle in the xy plane. In this case the light is said to be right circularly polarized. Figure 19.2 depicts a circularly polarized ray. If $E_x = E_y$ and $\phi_x - \phi_y = -90^\circ$, then the light is said to be left circularly polarized.

POLARIZATION PITFALL Not all books agree on the definition of right and left circular polarization. The definition presented here is consistent with the *OSA Handbook of Optics*: When the rotation of the electric field vector is clockwise with the observer looking opposite to the direction of propagation, the light is called *right circularly polarized*.

Figure 19.1

Vertically Polarized Wave:
 a. Three-Dimensional Perspective
 b. Looking Down the Propagation Axis, Projected onto the xy Plane
 Arrows Represent the Electric Field.
 Dashed Line Traces the Wave.
 Dots Mark the Amplitude



In the general case, where $E_x \neq E_y$ and $\phi_x \neq \phi_y$, the electric field vector traces an ellipse, and the electric field vector is said to be elliptically polarized.

Equation 19.1 may be written in complex notation, with the assumption that the real part of the equation should be taken:

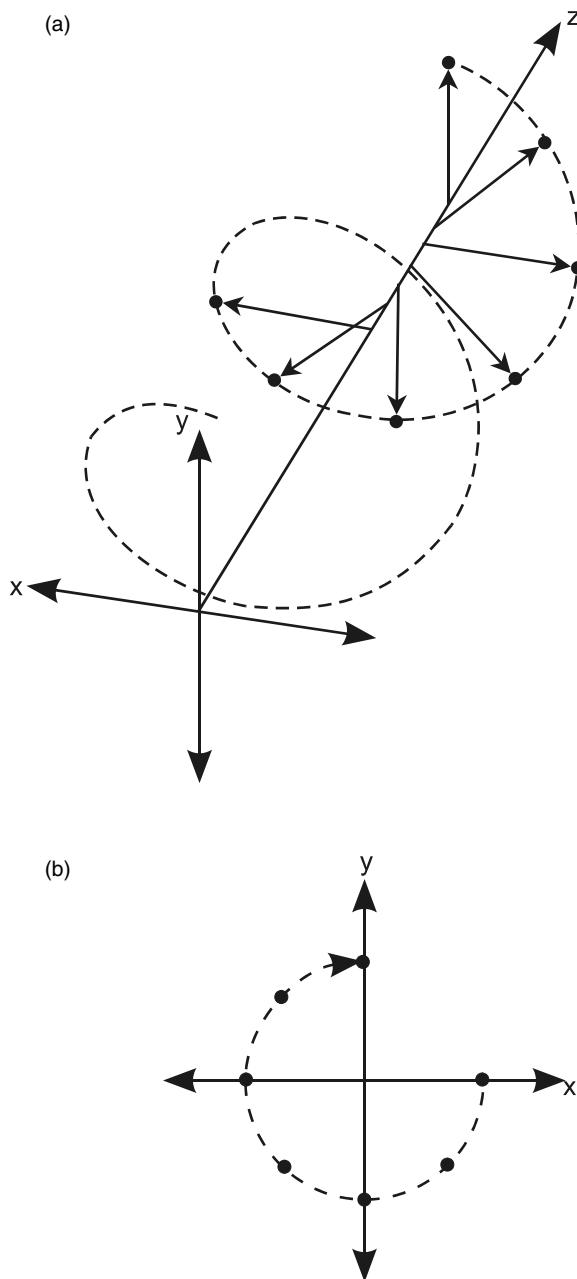
$$\vec{E} = \hat{x}E_x \exp i(kz - \omega t + \phi_x) + \hat{y}E_y \exp i(kz - \omega t + \phi_y) \quad (19.2)$$

A shorthand notation to express the polarization associated with this field is

$$\vec{J} = \begin{pmatrix} J_x \\ J_y \end{pmatrix} = \begin{pmatrix} E_x \exp i\phi_x \\ E_y \exp i\phi_y \end{pmatrix} \quad (19.3)$$

where it is understood that $\exp i(kz - \omega t)$ multiplies the vector to form the complex electric field.

Figure 19.2
Right Circularly Polarized Wave:
a. Three-Dimensional Perspective
b. Looking Down the Propagation Axis, Projected on the xy Plane
Arrows Represent the Electric Field.
Dashed Line Traces the Wave.
Dots Mark the Amplitude



In this notation the two-element column vector \vec{J} of Eq. 19.3 is called a *Jones vector*. The Jones vector is a simple way to express the polarization state of a fully polarized beam of light that is traveling in one direction. *The Jones vector is a convenient representation of the electric field that carries the minimum information necessary for describing polarization.*

Note that the power of the ray is proportional to the square of the amplitude, so the power P associated with a Jones vector \vec{J} is

$$P = J_x \cdot J_x^* + J_y \cdot J_y^* \quad (19.4)$$

where $*$ denotes the complex conjugate. Frequently the Jones vector is normalized to unity power for simplicity.

At a microscopic level the electric field is a physical quantity that always has a value at a particular point in time and space; however, it may not be stationary in time in space. In general the Jones vector magnitudes are time and space dependent:

$$\vec{J} = \begin{pmatrix} E_x(x,y,t) \exp(i\phi_x(x,y,t)) \\ E_y(x,y,t) \exp(i\phi_y(x,y,t)) \end{pmatrix} \quad (19.5)$$

If the values of E_x and E_y are not fully correlated, then the light is said to be partially polarized rather than fully polarized, and if there is no correlation the light is said to be unpolarized. In real life the observation of light takes place over a nonzero region of time and a nonzero volume of space. If the electric field direction changes over the spatial or temporal bandwidth of the detector, then the ray will appear to be less than fully polarized, even though its values are well-defined on a small scale. Thus, the polarization associated with light may depend on the spatial and temporal characteristics of the instrument that detects it.

Polarization Elements

A polarization element is any optical element with which the ray interacts, possibly with change of polarization state. For most optical elements (such as lenses, mirrors, polarizers, coated prisms, crystal retarders, etc) the incident and exiting electric fields are linearly related, and therefore the behavior of the element for a particular ray can be described by a 2×2 complex matrix \vec{J} , called a *Jones matrix*.

$$\vec{J}_{\text{OUT}} = \vec{J}_{\text{INC}} \cdot \vec{J} \quad (19.6)$$

where \vec{J}_{OUT} is the Jones vector of the exiting ray, and \vec{J}_{INC} is the Jones vector of the incident ray. The Jones matrix of an element with no effect on polarization is the identity matrix.

The usefulness of this Jones calculus lies in the fact that the polarization behavior of a series of optical elements can be described by the multiplication of the Jones matrices associated with each

$$\vec{J}_{\text{SYS}} = \vec{J}_Q \cdot \vec{J}_{Q-1} \cdot \vec{J}_2 \cdot \vec{J}_1 \quad (19.7)$$

where \vec{J}_i is the Jones matrix of the i th element and \vec{J}_{SYS} is the Jones matrix describing the sequence of Q elements.

POLARIZATION PITFALL Note that the Jones matrix of the first element the ray encounters lies the furthest to the right in the list.

Polarization elements may be classified in many ways. Two useful types of elements are polarization generators and polarization analyzers. Polarization generators prepare incident light of arbitrary polarization into a specific polarization state. Polarization analyzers determine how much of the incident light is in a particular polarization state. In analyzing the polarization behavior of an optical system it can be helpful to identify elements that function as polarization generators or as polarization analyzers.

Polarization Generators

Polarization generators are optical elements that produce light in a single polarization state, regardless of the incident polarization. For example, an element whose Jones matrix \vec{J}_{GEN} is given by

$$\vec{J}_{\text{GEN}} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \quad (19.8)$$

generates light polarized in the x direction regardless of incident polarization, because for an arbitrary incident Jones vector $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, the output Jones vector \vec{J}_{OUT} has only an x component:

$$\vec{J}_{\text{OUT}} = \vec{J}_{\text{GEN}} \cdot \vec{J}_{\text{INC}} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} a\alpha + b\beta \\ 0 \end{pmatrix} \quad (19.9)$$

Polarization Analyzers

Polarization analyzers are optical elements whose transmitted power is proportional to the power in one polarization state of the incident beam. For example, an element whose Jones matrix \vec{J}_{ANA} is given by

$$\vec{J}_{\text{ANA}} = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \quad (19.10)$$

where $|a|^2 + |b|^2 = 1$, analyzes the amount of the incident light that is in the x -polarization state: The output Jones vector \vec{J}_{OUT} is given by

$$\vec{J}_{\text{OUT}} = \vec{J}_{\text{ANA}} \cdot \vec{J}_{\text{INC}} = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha \cdot \begin{pmatrix} a \\ b \end{pmatrix} \quad (19.11)$$

and the power associated with the output ray is proportional to $\alpha \cdot \alpha^*$, which equals the power of the incident ray in the x polarization state.

Stokes Vectors

Stokes vectors provide an alternative description of polarized light based on measurement. Consider a beam of light propagating in the z direction and successively place polarization analyzers of known transmittance in the beam and measure the transmitted power. Define six polarization analyzers which respond, respectively, to horizontally polarized light, vertically polarized light, 45° polarized light, 135° polarized light, right circular polarized light, and left circular polarized light. When the readings are adjusted for the transmittances of the analyzers, the power of the incident light associated with each of the six polarization analyzers is determined: P_H , P_V , P_{45} , P_{135} , P_R , and P_L .

The four-element Stokes vector \vec{S} is then defined as:

$$\vec{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} P_H + P_V \\ P_H - P_V \\ P_{45} - P_{135} \\ P_R - P_L \end{pmatrix} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} \quad (19.12)$$

The top element of the Stokes vector represents the total power of the beam. The second element represents the degree to which the beam is

horizontally polarized; the third element represents the degree to which the polarization favors the 45° direction over the 135° direction; and the fourth element represents the degree to which the polarization favors right over left circularly polarized light. Some people prefer the notation with the S 's, and others prefer to use I , Q , U , and V .

A Stokes vector can describe unpolarized as well as polarized light. For an unpolarized beam the Stokes vector \vec{S}_U is

$$\vec{S}_U = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (19.13)$$

The degree of polarization (DOP) describes the extent to which the beam is polarized. The DOP associated with a Stokes vector \vec{S} is

$$\text{DOP} = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \quad (19.14)$$

For a fully polarized beam the DOP is one. For an unpolarized beam the DOP is zero. Any partially polarized beam may be represented uniquely as the sum of a fully polarized beam \vec{S}_P and an unpolarized beam \vec{S}_U . For an arbitrary beam with Stokes vector \vec{S} the decomposition is

$$\vec{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \vec{S}_P + \vec{S}_U = \begin{pmatrix} \sqrt{S_1^2 + S_2^2 + S_3^2} \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} + \begin{pmatrix} S_0 - \sqrt{S_1^2 + S_2^2 + S_3^2} \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (19.15)$$

Jones and Mueller Calculus

A polarization state may be represented as a Jones vector or as a Stokes vector. If the state is not fully polarized, then the Jones vector representation will have to carry the time or space variation explicitly, which is awkward and frequently unknown. The Stokes vector can more easily represent partially polarized light. The Stokes vector does not carry the absolute phase of the electric field, but for analyzing polarization behavior this is not a drawback.

The action of an optical element on a polarization vector can be described with a Jones matrix (transforming Jones vectors) or with a Mueller matrix (transforming Stokes vectors). The Jones matrix is a 2×2 complex valued matrix. The Mueller matrix is a 4×4 real-valued matrix. The Jones matrix \vec{J}_θ of an element rotated about the optical axis by angle θ is related to the unrotated Jones matrix \vec{J}_0 by a rotation transformation:

$$\vec{J}_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \cdot \vec{J}_0 \cdot \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (19.16)$$

The rotated form \vec{M}_θ of a Mueller matrix \vec{M}_0 is

$$\vec{M}_\theta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & -\sin 2\theta & 0 \\ 0 & \sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \vec{M}_0 \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (19.17)$$

Linear algebra teaches that two eigenvectors are associated with any 2×2 matrix, each with an associated eigenvalue. The physical state corresponding to an eigenvector of a polarization element is called an eigenpolarization. If a ray arrives at an element in an eigenpolarization, then the element does not change the polarization of the ray, although its average amplitude and phase may be affected. Polarization elements may be classified by their eigenpolarizations and eigenvalues.

Choosing the Jones Calculus or the Mueller Calculus

In analyzing the polarization behavior of an optical system, an optical engineer must determine whether to use the Jones Calculus or the Mueller Calculus (or neither). There is no hard-and-fast rule for this choice, but there is a rule-of-thumb:

When theoretically analyzing the passage of individual rays through a system in which only fully polarized light is present, use the Jones Calculus. The behavior of simple sequences of polarizers and retarders is usually most easily represented in the Jones Calculus. Polarization ray tracing software implicitly works using the Jones Calculus.

When describing measurements or when partially polarized light is present, use the Mueller Calculus.

The polarization vectors and matrices associated with simple polarization states and components are listed in Tables 19.1 and 19.2.

TABLE 19.1

Jones and Stokes vectors of selected polarization states

	Polarization State	Jones Vector	Stokes Vector
	Linear horizontal	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$
	Linear vertical	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$
	Linear 45°	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$
	Linear 135°	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$
	Linear θ	$\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$	$\begin{pmatrix} 1 \\ \cos 2\theta \\ \sin 2\theta \\ 0 \end{pmatrix}$
	Circular right	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$
	Circular left	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$

TABLE 19.2 Jones and Mueller matrices of selected polarization elements

Polarization Element	Jones Matrix	Mueller Matrix
Linear polarizer horizontal	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	$\frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Linear polarizer vertical	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$	$\frac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Linear polarizer θ	$\begin{pmatrix} \cos^2\theta & \cos\theta \sin\theta \\ \cos\theta \sin\theta & \sin^2\theta \end{pmatrix}$	$\frac{1}{2} \begin{pmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \cos 2\theta \sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta \sin 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Linear retarder horizontal, quarter wave	$\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$
Linear retarder horizontal, half wave	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$
Linear retarder horizontal, retardance δ	$\begin{pmatrix} \exp(i\delta/2) & 0 \\ 0 & \exp(-i\delta/2) \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\delta & \sin\delta \\ 0 & 0 & -\sin\delta & \cos\delta \end{pmatrix}$

Polarization Pitfalls of the Polarization Calculi

The simple form of the Jones and Stokes vectors makes them easy to manipulate but hides some underlying assumptions and facts:

1. A coordinate system underlies each polarization vector and matrix. As a ray transmits through an optical system it may be refracted and reflected many times. To describe the polarization of the ray,

the local coordinate system to which the polarization vector is referenced must change with each change of ray direction.

2. The polarization behavior of an optical element is likely to depend on wavelength, angle of incidence, and orientation of the plane of incidence.
3. The polarization behavior of an optical element may vary over its aperture.

Some Polarization Phenomena

This section presents a variety of polarization phenomena that may be encountered in an optical system. It is not an exhaustive list, merely one to demonstrate the range of polarization behavior that may arise in common optical systems.

Polarization of Uncoated Surfaces

The transmission and reflection of a ray of light at an optical interface depends on its polarization state. For an air-glass interface, for example, the *p* polarization component always transmits with higher efficiency than the *s* polarization component.

S and *p* refer to the linear polarization orientations that are perpendicular and parallel to the plane of incidence, respectively. (A useful memory aid is that the *s* component will skip along the surface and the *p* component will poke into it.) Because the plane of incidence depends on the orientation of the optical interface, the terms *s* polarization and *p* polarization are meaningless unless they are referred to a surface orientation.

POLARIZATION PITFALL A beam of light that is *s* polarized with respect to one surface may be *p* polarized with respect to the next surface, even if no change in electric field orientation occurs, because the plane of incidence of the second surface may be rotated with respect to the first. Also, a linearly polarized beam of finite aperture incident on a curved surface may be *s* polarized in some regions of the aperture, *p* polarized in other regions, and a mixture of *s* and *p* polarized in others.

For an air-to-glass interface the polarization behavior is described by Fresnel coefficients. The reflectance of glass at $n = 1.5$ is plotted in Fig. 19.3 for a range of angles of incidence for s and p polarization components. Note that the reflectance of the p polarization drops to zero at one angle, known as the *Brewster angle*. Even with unpolarized light incident, the reflection from a glass surface can be highly polarized.

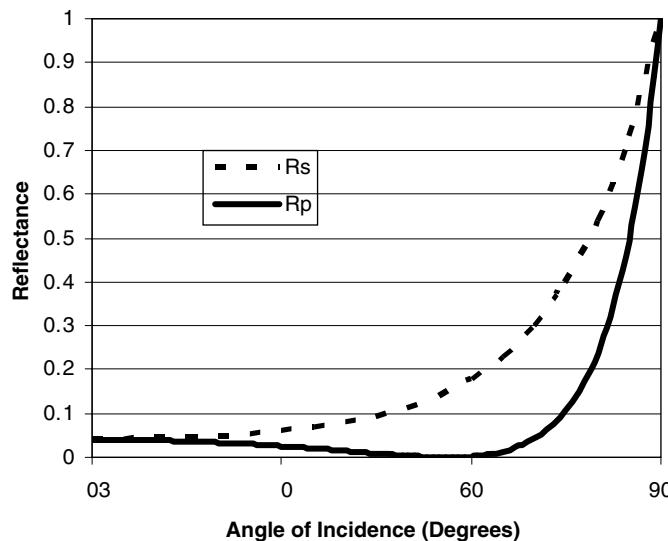
The Jones matrix governing the interaction in transmission is very simple when expressed in an $s-p$ coordinate system:

$$\vec{\vec{J}}_{\text{TRANS}} = \begin{pmatrix} t_s & 0 \\ 0 & t_p \end{pmatrix} \quad (19.18)$$

where $t_s = \sqrt{T_s}$ and $t_p = \sqrt{T_p}$ are the real-valued amplitude transmission coefficients for the s and p states, respectively. Unless the s and p polarization orientations coincide with the x and y directions, this matrix must be rotated into the ray's x y coordinate basis by a rotation as given in Eq. 19.16.

For air to glass or glass to glass interfaces the values of t_s and t_p are real. For metals (or generally, for media with complex refractive indices), the values of t_s and t_p may be complex.

Figure 19.3
Reflectance of s and p Polarized Light from an Air ($n = 1$) to Glass ($n = 1.5$) Interface. At an Angle of Incidence Equal to Brewster's Angle (56.3°) the p Reflectance Vanishes



Polarization of Coatings

Thin film coatings also produce transmission and reflection coefficients that are different for s and p polarized light. The coefficients may be complex, indicating that both the change in magnitude and the change in phase on interacting with a coating can depend on the incident polarization. The methods for calculation of the polarization-dependent transmission and reflection coefficients are given in many texts and software to perform these calculations either for an isolated surface or throughout an optical system are readily available.

Depending on the thin-film stack, a coating on glass may result in less or in more polarization when compared to an uncoated surface. As a rule of thumb, AR coatings tend to decrease the difference between t_s and t_p and between r_s and r_p at the interface.

Diattenuation and Polarizers

Diattenuation is the general phenomenon of polarization-dependent transmission. A diattenuator has one transmission coefficient for one eigenpolarization state and another transmission coefficient for the orthogonal state. Its Jones matrix \vec{J}_D can be expressed as

$$\vec{J}_D = \begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix} \quad (19.19)$$

where t_1 and t_2 are the real-valued amplitude transmission coefficients, and the Jones vector is expressed in the coordinate system corresponding to the eigenpolarizations of the element. In general this matrix needs to be rotated to the local coordinate system of the incident Jones vector in order to be applied.

The uncoated surface is a simple diattenuator. A diattenuator with much greater polarization effect is a linear polarizer, which, if aligned to pass x-polarized light, has a Jones matrix \vec{J}_{POL} , where

$$\vec{J}_{\text{POL}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (19.20)$$

If linearly polarized light is incident on an ideal linear polarizer, then the fraction of power T transmitted by the polarizer is

$$T = \cos^2(\theta) \quad (19.21)$$

where θ is the angle between the transmitting axis of the polarizer and the polarization orientation of the incident beam.

Linear polarizers come in several forms, including dichroic sheets, dielectric interfaces, wire grids, and crystalline prisms. Each has its pros and cons. There is no one *best* polarizer for all situations.

Retardance

Retardance is the phenomenon of polarization-dependent phase change. A retarder induces one phase change for one polarization state and another phase change for the orthogonal state. Its Jones vector \vec{J}_{RET} can be expressed as

$$\vec{J}_{\text{RET}} = \begin{pmatrix} \exp i\theta_1 & 0 \\ 0 & \exp i\theta_2 \end{pmatrix} \quad (19.22)$$

where θ_1 and θ_2 are the phase changes, and the Jones vector is expressed in the coordinate system corresponding to the eigenpolarizations of the element. In general this matrix needs to be rotated to the local coordinate system of the incident Jones vector in order to be applied.

Retardance occurs at coated surfaces, on reflection from tilted mirrors, on total internal reflection, and on propagation through birefringent media. A pure retarder has no effect on the transmitted power and cannot change the degree of polarization of the incident light. It can, however, transform one polarization state into another.

Sometimes retardance is stated in wavelengths. For a quarter wave retarder, $\theta_1 - \theta_2 = \pm 90^\circ$. For a half wave retarder, $\theta_1 - \theta_2 = \pm 180^\circ$. Most retarders are not achromatic; a quarter wave retarder at one wavelength could be a half-wave retarder at another. Stacks of multiple plate retarders can be designed in order to tailor the wavelength or angle of incidence properties.

POLARIZATION PITFALL Vendors may supply a retarder with its slow or fast axis orientation marked, but they don't always get it right.

Birefringence

Birefringence is one physical property that can lead to the optical phenomenon of retardance. When light is incident onto a birefringent medium, one incident polarization state experiences one refractive index

in the birefringent medium, and the orthogonal polarization sees a different refractive index. This can result in splitting the ray into two transmitting paths, and the phase of each polarization state accumulating at a different rate. The physics of birefringent media are described in many texts. Here we make only a few comments corresponding to uniaxial birefringent media, such as a calcite crystal.

A description of a uniaxial medium will include its surface orientations, two refractive indices n_o (ordinary) and n_e (extraordinary), and the orientation of the crystal axis. Light inside the medium that is polarized in the direction orthogonal to the crystal axis will see the ordinary refractive index. Light inside the medium that is polarized orthogonal to the ordinary mode will see a refractive index between n_o and n_e , given by $n(\theta)$,

$$\frac{1}{n^2(\theta)} = \frac{\cos^2(\theta)}{n_o^2} + \frac{\sin^2(\theta)}{n_e^2}, \quad (19.23)$$

where θ is the angle between the direction of wave propagation and the crystal axis.

POLARIZATION PITFALL Note that the birefringent medium is characterized by two refractive indices, n_o and n_e , but the refractive index $n(\theta)$ of the extraordinary mode may take on any value between n_o and n_e .

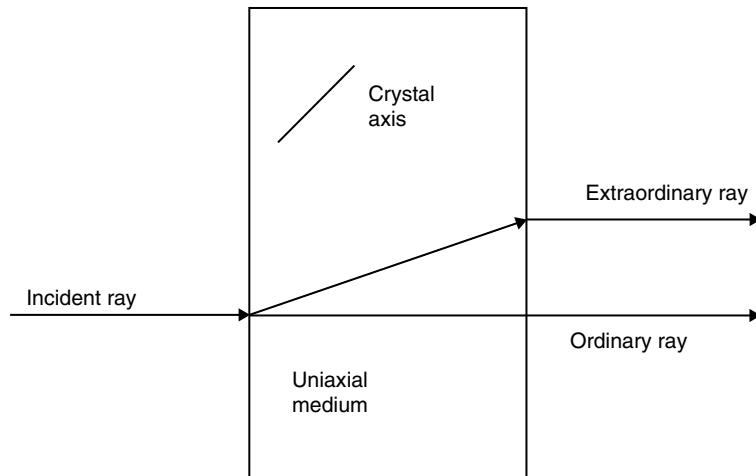
For light normally incident onto a uniaxial crystal whose crystal axis lies in the xy plane, the ray will continue to propagate straight through the crystal, but the phase accumulation of one polarization component will be different than the phase accumulation of the orthogonal component. The ray will exit the element having experienced retardance R given by

$$R = \frac{2\pi d(n_e - n_o)}{\lambda} \quad (19.24)$$

where d is the thickness of the crystal.

For light incident at nonnormal incidence the ray will split into two paths which separate in space. A further oddity of birefringent media is that the wave propagation direction (normal to the wavefront) does not necessarily coincide with the direction of the energy propagation (ray direction). In an element called a walkoff plate, the crystal axis is partially tilted into the plane of the entrance surface. A beam at normal incidence will split into one *ordinary* beam that is undeviated and into one *extraordinary* beam whose ray vector tilts away from the optics axis.

Figure 19.4
Birefringent Walkoff Plate: On Entering the Crystal the Incident Ray Splits into Two. The Ordinary Ray Continues without Deviation. The Extraordinary Ray "Walks Away" from the Optical Axis While Inside the Crystal. (The Deviation Angle is Exaggerated for Clarity)



Nonetheless, Snell's law still holds, and the wavefront remains parallel to the xy plane. Figure 19.4 depicts a walkoff plate with an exaggerated walkoff angle.

Note that for most birefringent optical materials, $n_e - n_o$ is quite small, often much less than 0.01, so the beam separation occurring on transmission through a crystalline retarder, for example, may be negligible. However, applications exist where this phenomenon is central to the function of the element, such as walkoff plates in fiber optic isolators and in birefringent blur filters.

Liquid crystals are liquid collections of birefringent molecules whose alignment, and therefore retardance, may be controlled by an external electric field. When polarizers are placed around a cell containing appropriate liquid crystal materials with particular thicknesses and applied electric fields, pixels of variable optical transmission may be constructed. This is the principle behind LCD televisions and liquid crystal based projection displays.

Stress Birefringence

An unstressed optical glass may become birefringent when stress is placed on it. This can occur due to mechanical pressures applied by a mount or to thermally induced stresses resulting from temperature gradients inside the glass. The susceptibility of a glass to stress is given by its stress

optic coefficient, which is found in glass catalogs. This relates the refractive index change (birefringence) induced by an amount of stress. Stress birefringence can be greatly reduced by using glasses such as SF57, which has a very low stress optic coefficient, but these glasses typically have other substantial drawbacks which may preclude their use. Molded plastic components are particularly susceptible to stress birefringence unless great care is taken in their manufacture.

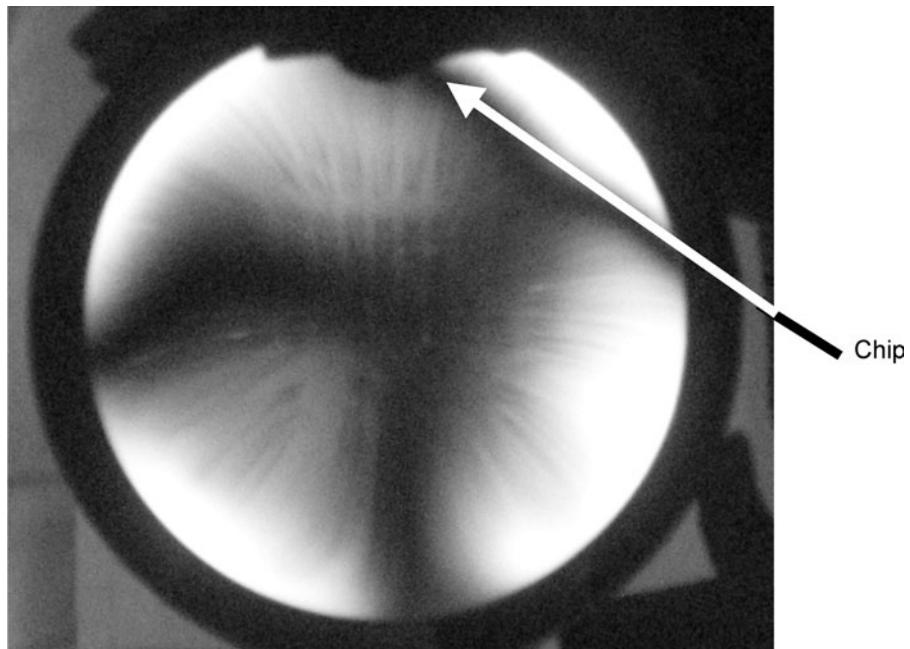
A thermal/structural/optical analysis of a system can be performed to model the effects of stress birefringence. The thermal model calculates the thermal profile of the optics and their mounts. The structural model calculates the resulting stresses throughout the volume of the optical element. The optical analysis converts these stresses into birefringence maps. Finally, the retardance accumulation is calculated on a ray-by-ray basis through the nonuniform medium. Needless to say, this process can be quite complicated. Avoidance and control of stress birefringence is frequently an important goal in the design of polarization critical optical systems. For example, a radial force on an optical lens of a pound can produce a retardance in the lens greater than 0.1 waves, dramatically altering the polarization state of the light passing through it.

Stress birefringence can be observed by viewing the object under test between crossed polarizers as shown in Fig. 19.5. The leakage pattern indicates areas of polarization change resulting from the stress. The pattern may somewhat resemble an interferogram. However, the pattern represents a phase difference map between the orthogonally polarized wavefronts, not the wavefront variation that is common to both polarizations.

Interpretation of the patterns of colored fringes seen in white light for a transparent object between crossed polarizers is quite complex because they represent a convolution of the stress and the direction of stress. For example, an object with a large stress-induced birefringence may show no transmission variation at all if the stress direction lies along the transmission direction of one of the linear polarizers. Also, if the direction of stress varies over the object, but the magnitude of that stress remains constant, then there will be a varying fringe pattern that could be misinterpreted as a variation in the magnitude of the stress across the object. Some of this confusion can be sorted out by using orthogonal circular polarizers instead of linear polarizers. Then the transmission is independent of the stress direction. However, most circular polarizers are only circular at a single wavelength and if we use monochromatic illumination we lose important clues to the magnitude of stress provided by the fringe colors. Observing stressed objects

Figure 19.5

Variable Stress across This Lens Produces a Transmission Pattern When Viewed between Orthogonal Polarizers



between polarizers is an important testing method for the commercial glass industry. Glass bottle and vial manufacturers use this method for process monitoring.

Depolarization

Depolarization is the conversion of polarized light into unpolarized light.

POLARIZATION PITFALL Some authors confusingly refer to depolarization as any change in polarization away from the desired polarization.

POLARIZATION PITFALL A depolarizer does not necessarily change the degree of polarization of an incident beam, and an element that changes the degree of polarization is not necessarily a depolarizer.

A depolarizer adds randomness to the electric field vector, causing the components of the Jones vector to vary in a manner that is fast compared to the detection bandwidth. A spatial depolarizer varies the polarization properties over the beam aperture on a spatial scale too small for the detector to distinguish. A temporal depolarizer dynamically varies

its polarization properties in time at a rate faster than the detector bandwidth can follow. A wavelength depolarizer statically varies its polarization properties over a wavelength bandwidth small with respect to the wavelength range of the detection system. None of these methods really depolarize light in the strict sense. A better name for them might be polarization scramblers.

Note that a depolarizer cannot be expressed as a Jones matrix, because the Jones matrix represents the conversion of one fully polarized state into another. The Mueller matrix of a perfect, complete depolarizer \vec{S}_{DP} is

$$\vec{S}_{DP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (19.25)$$

Depolarizers may, however, have more complex Mueller matrices, with the amount of depolarization depending on the incident polarization state.

A depolarizing element may be a nuisance in a system that depends on maintenance of a fixed, uniform polarization state. On the other hand, a depolarizer may be introduced into a radiometer in order to reduce the polarization dependence of its responsivity.

Blackbodies

Blackbody radiation is often thought of as incoherent and completely random, but the radiation emitted by a heated smooth surface will be polarized, preferentially in the p orientation. As a result, the light emitted from, for example, a filament lamp can have a small degree of polarization, even if the intervening optics are nonpolarizing.

Maltese Cross

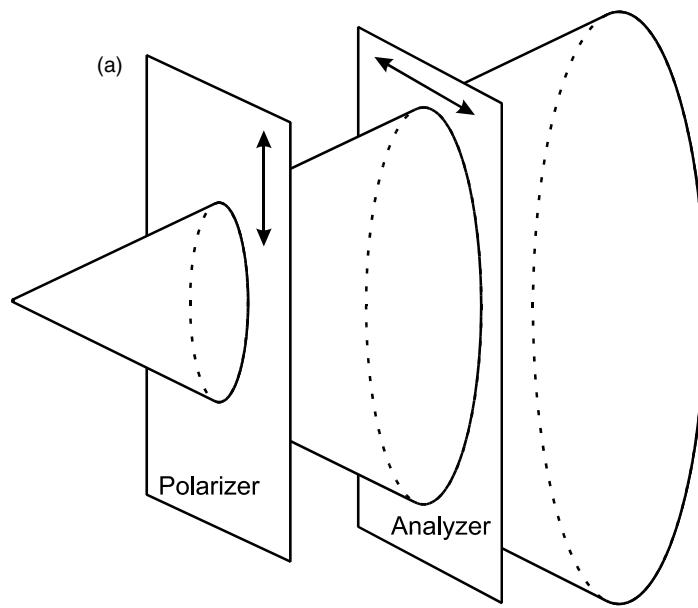
Two *good* high-extinction linear sheet polarizers whose absorption axes are mutually perpendicular transmit very little of the incident power. However, when placed in a converging beam of light, the transmittance is nonzero. Upon examination of the leakage pattern in angular space, one sees that the transmittance is near zero at normal incidence and

along the axes parallel to the first or second polarizer absorption axes. However, for skew rays, leakage appears that grows with the angle of incidence. The shape of the leakage pattern resembles a Maltese Cross, as shown in Fig. 19.6.

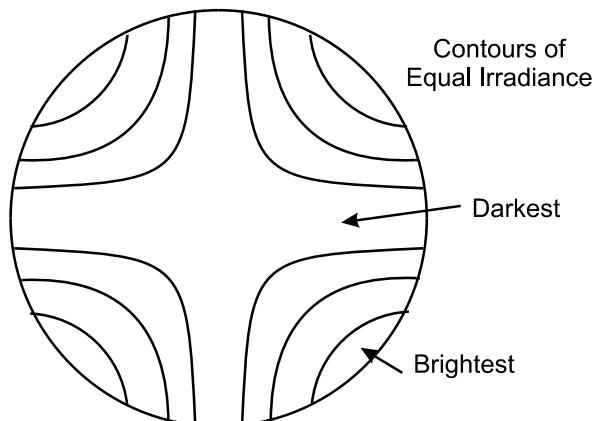
This effect appears even when extinction ratio of the polarizers is very high and when the polarizers are specified to have a large field of view. Typically when vendors speak of a field of view for a polarizer,

Figure 19.6

- a. Propagation of a Cone of Light through Crossed Linear Polarizers
b. The Irradiance Pattern of the Leaked Light Resembles a Maltese Cross



(b)



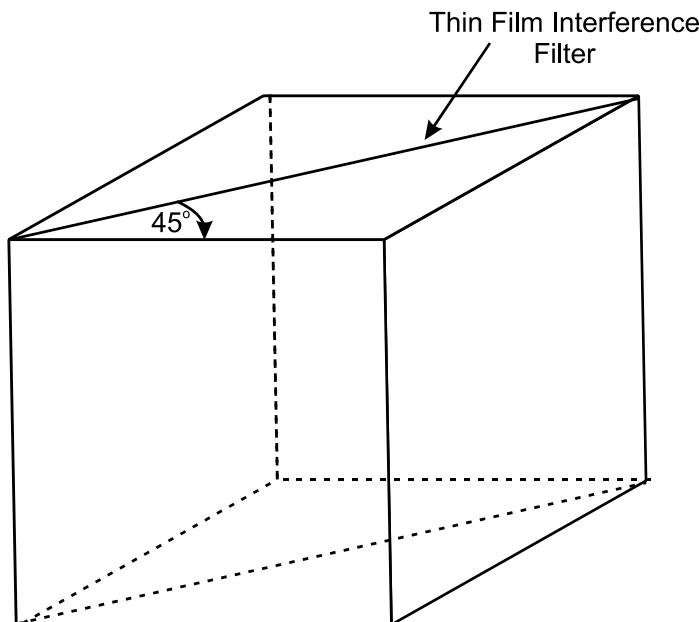
they refer to the angles over which the transmitted ray will be highly polarized for unpolarized light incident. However, the orientational variation of the polarization state generated by the first polarizer may fail to align perfectly to the absorption axis of the second in these diagonal angular regions. The leakage due to the Maltese Cross effect depends strongly on the details of the polarizer. Many sheet polarizers are laminated in cellulose acetate butyrate to protect the rather delicate polarizing layer. This protective material can exhibit substantial retardation when viewed at a large angle of incidence and this further degrades the transmittance between crossed polarizers at large angles of incidence.

Polarizing and Nonpolarizing Beamsplitters

Beamsplitter cubes are made to split and recombine beams of light. As shown in Fig. 19.7, a multilayer coating on the interior diagonal of a cube of glass determines the polarization characteristics. A polarizing beamsplitter transmits the p polarization state and reflects the s polarization. The coating prescription determines the extinction ratio of the polarizer, the transmission efficiency, wavelength dependence, and the angular dependence of the device.

Figure 19.7

Polarizing Beam Splitter Cube A Thin-Film Interference Filter on the Inner Diagonal of a Cube of Glass Is Optimized to Transmit p Polarized Light Only



Because the *s* and *p* orientations depend on the plane of incidence, these polarizers tend to be most useful over a small angular range around normal incidence.

From the name “nonpolarizing beamsplitter” one might assume that this is an element whose behavior is independent of polarization state. However, that is not the case. Most of these elements have been designed to provide 50% transmission and 50% reflection of incident unpolarized light, but the polarization state of the transmitted or reflected beam is unspecified. These elements are rarely useful in a polarization critical system.

Scattering and Integrating Spheres

The near-Lambertian optical surfaces made from pressed PTFE (for example, Spectralon) can partially depolarize incident light. The scattering from the rough surface and from the volume scattering centers tends to randomize the polarization state. Nonetheless, at the near-specular directions at high angles of incidence even Spectralon is not a very good depolarizer, retaining some of the preference for scattering *s* polarization in the specular direction. When Spectralon is used to line the interior surface of an integrating sphere, however, the multiple reflections (with a wide range of angles and orientations) ensure that the radiance exiting the sphere is very close to unpolarized and independent of the polarization entering the sphere. (This assumes that the entrance and exit apertures are suitably small with respect to the surface area of the sphere.)

High NA Systems

At high numerical aperture (NA) the approximations underlying geometric optics begin to break down, and the analysis of polarization behavior becomes more difficult. Treatment of polarization in this regime is beyond the scope of this chapter. Such high NA systems are sometimes encountered in photolithography and optical recording, for example.

Polarization Control Nuts and Bolts

Polarizers and retarders or waveplates are the optical nuts and bolts we can use for controlling the polarization of light. They come in many varieties and this section should help you decide what kind is best for your application.

Dichroic Sheet Polarizers

The most common polarizer type is the dichroic sheet polarizer. It consists of a matrix of oriented dye molecules or crystals on a polymer substrate. Often the dye is iodine and the polymer is polyvinyl alcohol that has been stretched longitudinally to orient the dye molecules. These molecules preferentially absorb light polarized along the long axis of the molecule. As explained earlier the dye matrix is laminated in a protective material, usually a plastic or glass. This is the polarizer for computer screens and video screens and is available in sizes greater than a square meter. Its large clear aperture and low cost are its greatest benefits. It has limited damage threshold because the light of the rejected polarization direction is absorbed by the polarizer. A typical power limit is one watt/cm² and, of course, high temperatures can melt the polarizer.

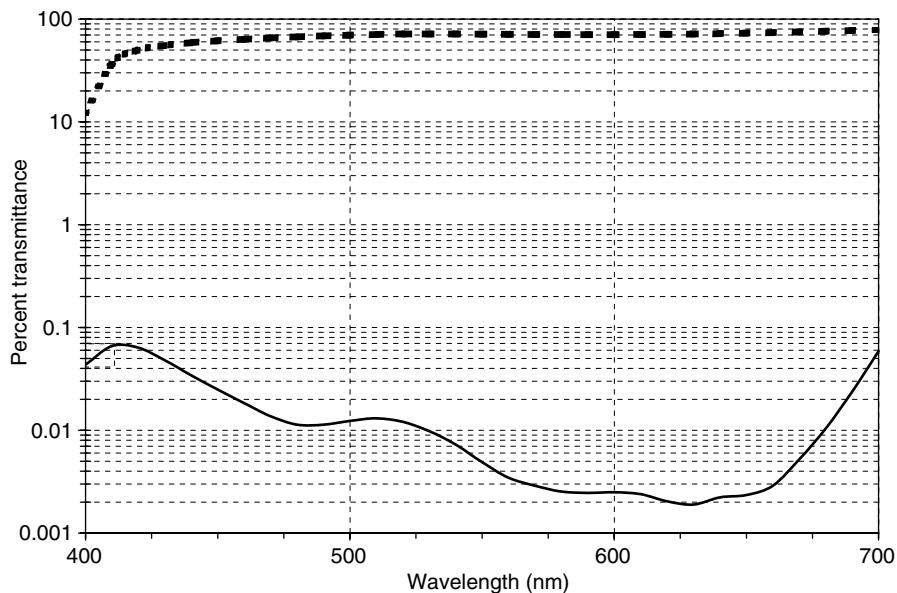
The dyes have a limited wavelength range for good performance and tend to fall into two categories. One range is from about 400 to 750 nm. These polarizers absorb all light at wavelengths shorter wavelengths than 400 nm and pass all polarizations at wavelengths longer than 750 nm as shown in Fig. 19.8. The second range is from about 800 to 2000 nm. Most light is absorbed below 800 nm and light of all polarizations is passed above 2000 nm.

The extinction ratio between crossed dichroic sheet polarizers is excellent, which often surprises engineers who expect low cost to mean poor performance. The transmitted wavefront distortion and reflection losses can be low if the polarizers are glass laminated although the cost of glass-mounted polarizers is much higher.

Metal Wire Polarizers

These polarizers have an array of conductive metal wires or whiskers either deposited on one surface or imbedded into a glass substrate. Usually

Figure 19.8
Performance of a
Dichroic Sheet Polarizer. The Upper,
Dashed, Curve Is for a Pair of Parallel
Polarizers and the
Lower, Solid, Curve Is
for a Pair of Crossed
Polarizers

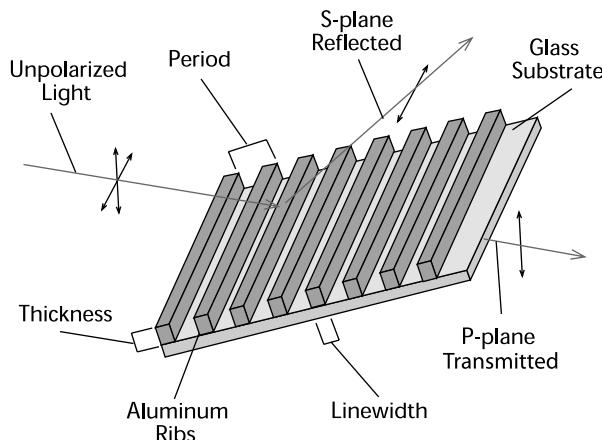


the metal is gold, aluminum, or copper and always the wires or whiskers have an axis orientation that is common across the aperture. Light polarized along this axis is rejected and light polarized perpendicular to this axis is passed. The rejected light is specularly reflected if the metal wires are deposited on one face of a polished glass substrate as shown in Fig. 19.9. The wavelength range of best performance depends on which metal is used and on the spacing of the wires. For best performance the wire spacing should be small compared to the wavelength of light to be polarized. Figure 19.10 shows the performance of one type of metal wire polarizer marketed under the name VersaLight. This polarizer has aluminum wires photolithographically patterned onto one surface of a glass substrate. The wire spacing is less than 100 nm and the rejected light is specularly reflected with good polarization fidelity. This means that the polarizer acts like a polarizing mirror even at normal incidence with the polarization direction in the reflected beam the same as the wire direction.

Pairs of these polarizers show extinction ratios (parallel to crossed transmissions) that range from less than 10:1 to greater than 10,000:1

Figure 19.9

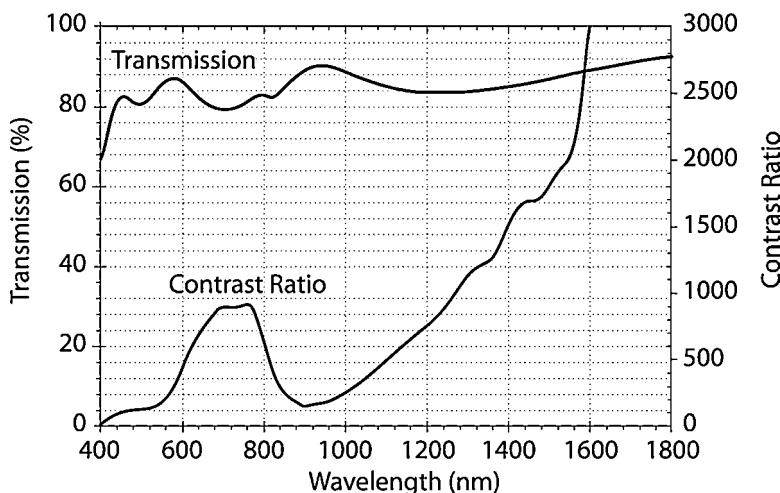
A Schematic Diagram of the Surface of a Metal Wire Polarizer. The Period of the Wires Must be Small Compared to the Wavelength of Light



depending on the wavelength, metal type, and wire spacing and geometry. Some types of these polarizers work well over a large wavelength range of 450 nm to more than 2000 nm. They can withstand temperatures of several hundred degrees Celsius without damage and power densities greater than 50 kW/cm^2 . Costs are much higher than for dichroic sheet polarizers.

Figure 19.10

Performance of a VersaLight Metal Wire Polarizer



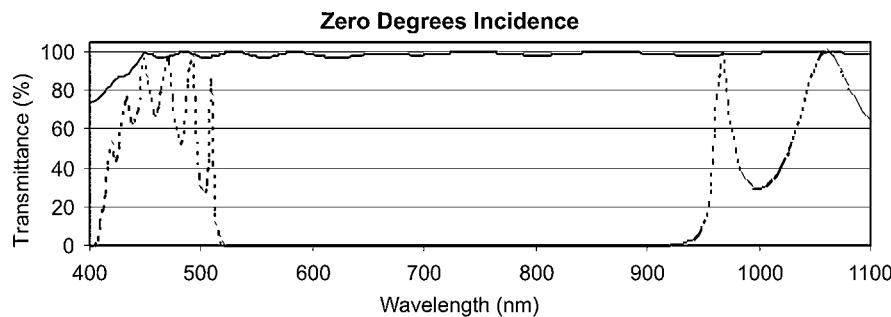
Polarizing Beamsplitter Cubes

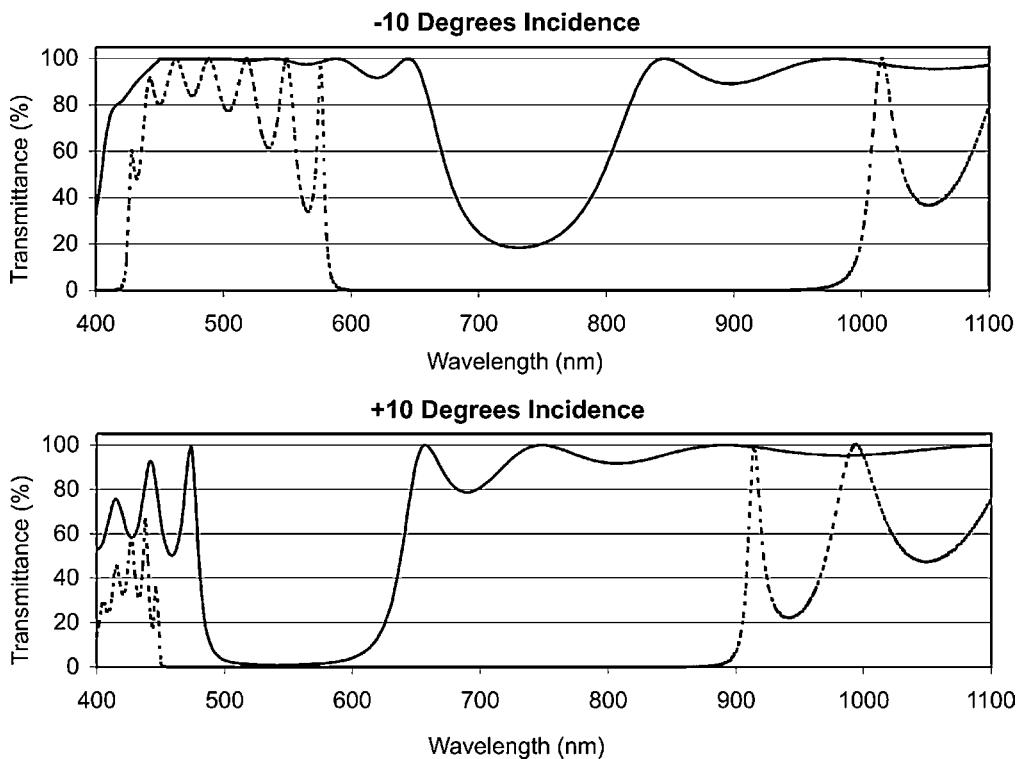
These cubes are made of two right angle prisms with some material on the hypotenuse to polarize the incident light. One linear polarization is transmitted and the orthogonal linear polarization is reflected. Most commonly a thin film stack of alternating high- and low-index materials of quarter wave optical thickness is coated on the hypotenuse of one prism as shown in Fig. 19.6. This acts like a pile-of-plates polarizer with performance enhanced by interference effects. Usually these thin film stacks are variants of the MacNeille design and always they reflect *s* polarized light and transmit *p* polarized light. The wavelength range depends upon the design of the thin film stack and the index of refraction of the prisms.

Figure 19.11 shows the performance for a MacNeille cube that works well over the wavelength range of 550 to 900 nm. These cubes will tolerate relatively high flux levels of 500 watts/cm². Usually the adhesive joining the two prisms breaks down before the coating. Optical contacting improves the damage threshold. The angular field of view in these polarizers is smaller than for all other types of polarizers described in this chapter. Figure 19.12 shows the change in performance of a MacNeille cube for rays for a 10° change in angle of incidence, in the plane of incidence. The angular field is further limited for skew rays because the plane of incidence rotates about the optical axis for these rays and the *s* and *p* polarization directions are defined by this plane. The contrast ratio in the transmitted beam on these cubes is usually better than 1000:1 but the reflected contrast ratio is lower, often less than 100:1. The reflected contrast ratio can be improved by adding a *cleanup* dichroic sheet polarizer downstream in the reflected *s* polarized beam.

Figure 19.11

Performance of a Thin Film Polarizing Beamsplitter Cube at Normal Incidence. The Upper, Solid, Curve Is the Transmission for *p* Polarized Light and the Lower, Dashed, Curve Is for *s* Polarized Light



**Figure 19.12**

Off Normal Performance of a Thin Film Polarizing Beamsplitter Cube. The Upper, Solid, Curve is the Transmission for p Polarized Light and the Lower, Dashed, Curve Is for s Polarized Light

POLARIZATION PITFALL Texts and polarizer manufacturers use the terms contrast ratio and extinction ratio interchangeably to describe the purity of polarization produced. Sometimes this is meant to be the ratio of parallel polarizer to crossed polarizer transmission measured against another like polarizer and sometimes it is against a perfectly polarized beam. This contrast or extinction ratio is dependent on many factors that are important and often not given. These include f ratio of the test beam and the solid angle of the detector as viewed from the polarizer surface.

Another type of polarizing beamsplitter cube replaces the thin film stack with a metal wire polarizer layer on the hypotenuse. These have a much better angular field of view and wavelength range than the MacNeille cube as shown in the measurements of contrast ratio listed in Table 19.3.

TABLE 19.3

Contrast Ratio
Measured Against
a Nearly Perfect
Linearly Polarized
Collimated Beam

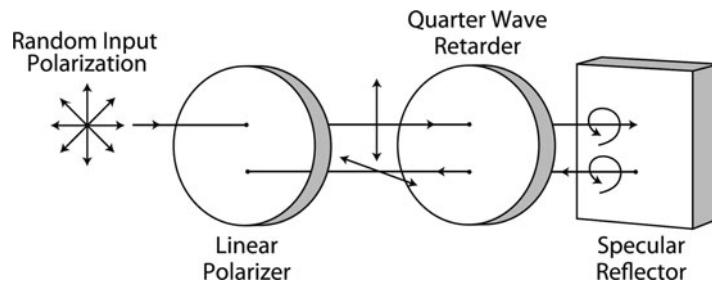
Angle of Incidence	Wavelength			
	450 nm	649 nm	1009 nm	1550 nm
-30°	29	490	1,000	4,100
-20°	29	570	1,100	2,000
-10°	29	580	1,400	4,000
0°	36	680	2,000	8,300
+10°	39	790	1,400	18,000
+20°	36	760	1,300	20,000
+30°	43	760	3,200	67,000

This cube absorbs approximately 20% of the incident light and consequently has a lower damage threshold than the MacNeille cube. The direction of polarization is defined by the wire direction, not by the plane of incidence. This means that the skew rays have the same polarization direction as those in the plane of incidence defined by the ray normal to the cube face. Also, since the wire direction defines the polarization direction it is no longer necessary for the reflected beam to be *s* polarized and the transmitted beam *p* polarized, as it is in the data in the above table. In fact cubes can be built with any orientation for the two orthogonally polarized beams including an *s* transmitted and *p* reflected configuration. This latter configuration has a lower contrast ratio for both beams.

Circular Polarizers

Although most polarizers are linear, there are applications where circular polarizers are needed. One example is in the stress analysis in glass bottles as discussed in “Stress Birefringence” section. Another example is in isolating lasers from specular back reflections as shown in Fig. 19.13. The most common type of circular polarizer combines a quarter wave retarder with a linear polarizer. The linear polarizer transmission axis

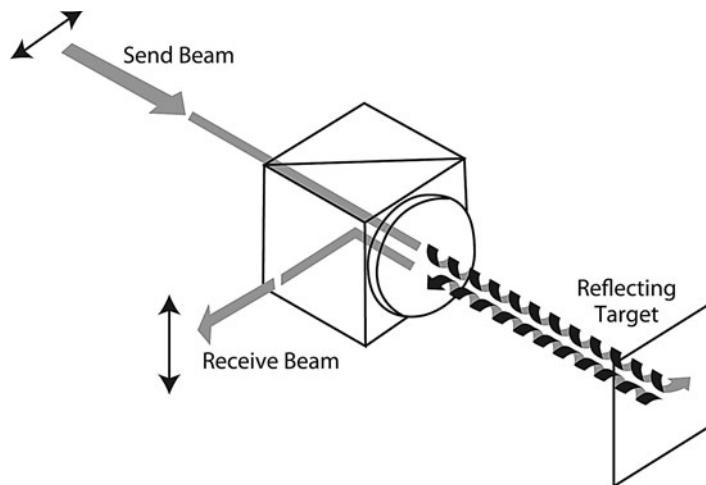
Figure 19.13
Specular Reflection Reverses the Handedness of a Circularly Polarized Beam. This Reflected Beam Becomes Horizontally Polarized on the Second Pass through the Quarter Wave Retarder and Is Blocked by the Vertical Linear Polarizer



is at 45° to the fast and slow axes of the following quarter wave retarder. These circular polarizers work well at only one wavelength unless the quarter wave retarder is achromatic. If the linear polarizer is a polarizing beamsplitting cube, the circular polarizer is sometimes called a *beam separator* because of the separation of the outbound and reflected beam in the configuration shown in Fig. 19.14.

A second type of circular polarizer uses a nematic liquid crystal layer to polarize light by Bragg reflection. The birefringent liquid crystal molecules are arranged in helices with the helix axis perpendicular to the polarizer face. Light of one circular polarization is passed and the orthogonal circular polarization is reflected. They work well for wavelengths of light close to the pitch length of the helix. Usually the contrast

Figure 19.14
The Beam Separator Separates the Outbound and Specularly Reflected Return Beam. It Is a Thin Film Beamsplitting Polarizer Cube with a Quarter Wave Retarder Attached to the Exit Face. The Retarder Fast Axis is at 45° to the Polarization Directions



ratio is above 1000 for a wavelength range of 50 to 100 nm. The liquid crystal layer can be crosslinked into a polymer in a manner that produces a continuously variable pitch length through the liquid crystal. The variable pitch length broadens the wavelength range so that good performance is possible over the entire visible wavelength range. This type of circular polarizer is not commonly available commercially.

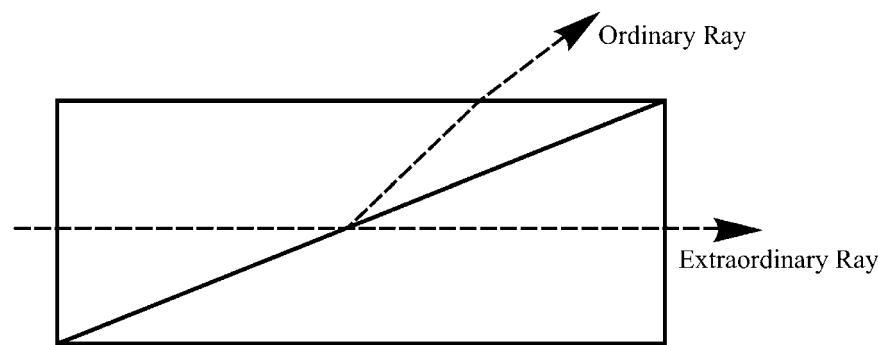
Birefringent Polarizers

There is a class of polarizers that use the birefringence of uniaxial crystals for polarization selection. High birefringence crystals perform best in these polarizers and the most commonly used crystal is calcite, which has a birefringence of about 0.17 for visible light. The simplest of these is the walkoff plate already shown in Fig. 19.4. This plate separates the unpolarized incident beam into two spatially separated beams of orthogonal linear polarizations. The spacing of the beams increases with the birefringence and with the thickness of the plate.

There are several other polarizers that play on this theme of different indices for different polarization directions. One of the more common ones is the Glan–Thompson polarizer shown in Fig. 19.15. The extraordinary ray is reflected at the cemented interface between two calcite prisms and the ordinary ray is transmitted. The extraordinary ray sees a

Figure 19.15

The Glan-Thompson Calcite Prism Polarizer Passes the Extraordinary Ray and Reflects the Ordinary Ray by Total Internal Reflection. The Optic Axis of the Calcite is Perpendicular to the Page and Parallel to the Polished Faces of the Calcite Prisms. The Transmitted Polarization Is in the Direction of the Optic Axis



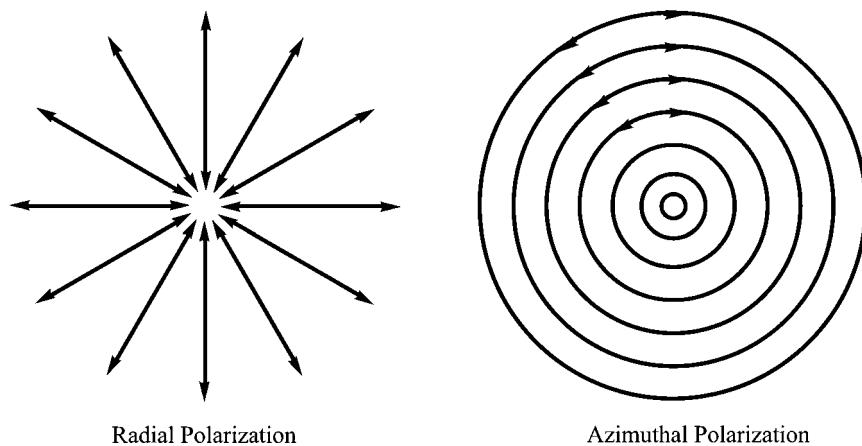
higher index of refraction and is totally internally reflected at the interface. The ordinary ray sees a lower index in the calcite and can pass through the prism interface. The crystal optic axis is parallel to all polished faces of the prisms.

These can have an excellent angular field of view and usually have an excellent contrast ratio greater than 100,000. The wavelength range can be as large as 230 to 2500 nm, which is the range for which calcite transmits. They are expensive and limited to apertures smaller than about 30 mm in diameter in part because of the limited availability of optical quality calcite. This crystal is mined and has never been grown in commercially significant quantities and sizes. The crystal barium borate (BBO) is replacing calcite in some birefringent polarizer applications. Both calcite and BBO are soft and more difficult to polish than most common optical glasses.

Patterned Polarizers

A printing process can make dichroic sheet polarizers with a spatially dependent polarization direction. This is used to make display signs that show motion when a second polarizer is rotated in front of the patterned one. There are emerging technical applications for spatially continuously variable polarizers, but methods for making these are beyond the scope of this chapter and these polarizers are not yet readily available commercially. They can be used to produce both radially polarized light and azimuthally polarized light as shown in Fig. 19.16.

Figure 19.16
Radially and
Azimuthally Polarized
Light



Switchable Polarizers

Some specialized polarizers can be switched on and off. In one state they pass light of all polarizations and in the second state they pass only one polarization. A pair of orthogonal switchable polarizers will pass nearly 100% of incident light in the open state, which is much better than the typical maximum transmission of about 40% for a pair of ordinary polarizers. One configuration for the switchable polarizer is a guest-host mixture of a liquid crystal and dichroic polarizer dye molecules. The dye molecules follow along when the liquid crystal molecules reorient under an applied electric field. When these molecules have their long axis parallel to the light path they do not polarize and when their long axis is perpendicular to the light path they do polarize.

Retarders

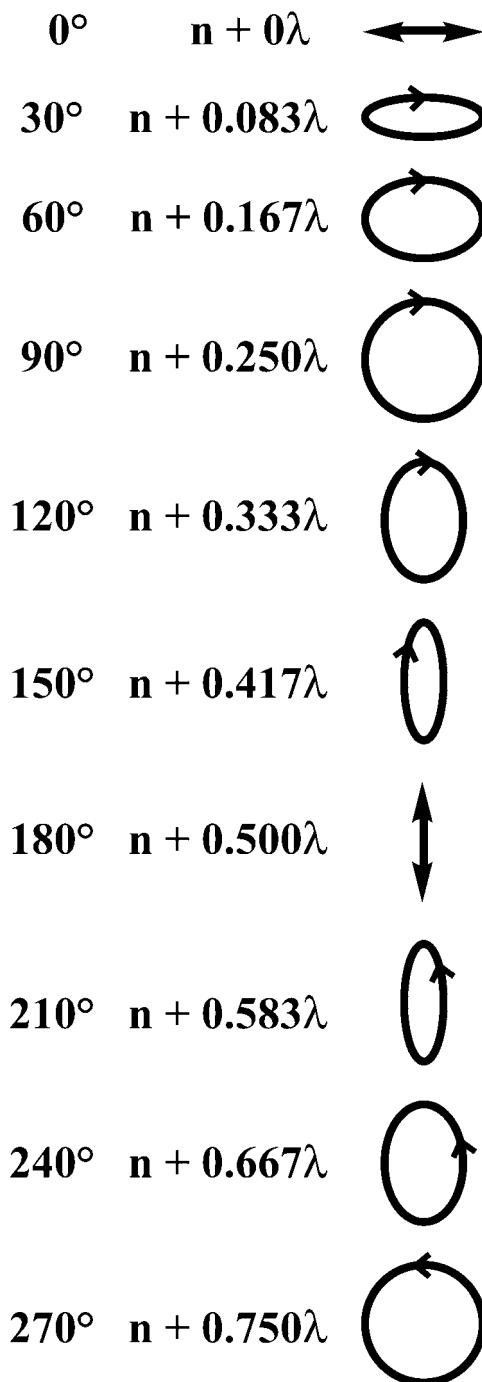
Retarders are polarization modifiers. They do not polarize unpolarized light and they do not change the degree of polarization in a partially polarized beam but they usually will change the polarization state of an incident polarized beam. There are many choices to make when selecting a retarder. First, of course, is the desired retardation. Beyond that is a matrix of parameters such as field of view, temperature range, damage threshold, and wavelength range. Not all of these are maximized with the same retarder type and material.

As discussed in “Retardance” section, the most commonly used retarders are half and quarter wave retarders. A quarter wave retarder can convert linearly polarized light to circularly polarized light and will convert circularly polarized light to linearly polarized light. A half-wave retarder can rotate a plane of linear polarization and can reverse the handedness of a circularly polarized beam. Use the Mueller or Jones matrices in Table 19.2 together with the rotation matrices in Eq. 19.16 to compute the polarization change for other retarder and input polarization configurations. Figure 19.17 shows how the output polarization form changes for different retardation values when the input polarization is linear and the retarder fast and slow axes are 45° to the polarization direction.

All common retarders are linear retarders, which means that they preferentially retard the phase of one state of linear polarization relative to that of the orthogonal linear polarization. There are also circular retarders but we will not discuss these here.

Figure 19.17

The Output Polarization Form for Horizontal Linearly Polarized Input to an Electrically Variable Retarder. The Retarder Optic Axis Is at 45° to Horizontal



Usually the retardation of phase results from transmission through a birefringent material as described earlier in “Birefringence” section. Retarders of this type are often called waveplates. The function of these waveplates is described in “Retardance” section. The material can be a crystal such as quartz, calcite or mica, but newer lower cost and more versatile liquid crystal and polymer materials are replacing crystals in many applications.

Retarder Mathematics

Retardance is most commonly stated in waves because this describes best how the retarder will modify the input polarization. However, for this to be meaningful we must also state the wavelength of the light. A retarder that is half wave at a wavelength of 400 nm will only be approximately a quarter wave retarder at 800 nm. This is because retardation is a distance, so a half wave retarder at 400 nm can also be said to retard the slow component of the wave by 200 nm. The distance in length units by which a retarder delays the slow component changes slightly with wavelength. This is because the birefringence is a slow function of wavelength, much like the index of refraction of a glass. This is why the half-wave retarder at 400 nm is only approximately quarter wave at 800 nm.

We can restate Eq. 19.24 as

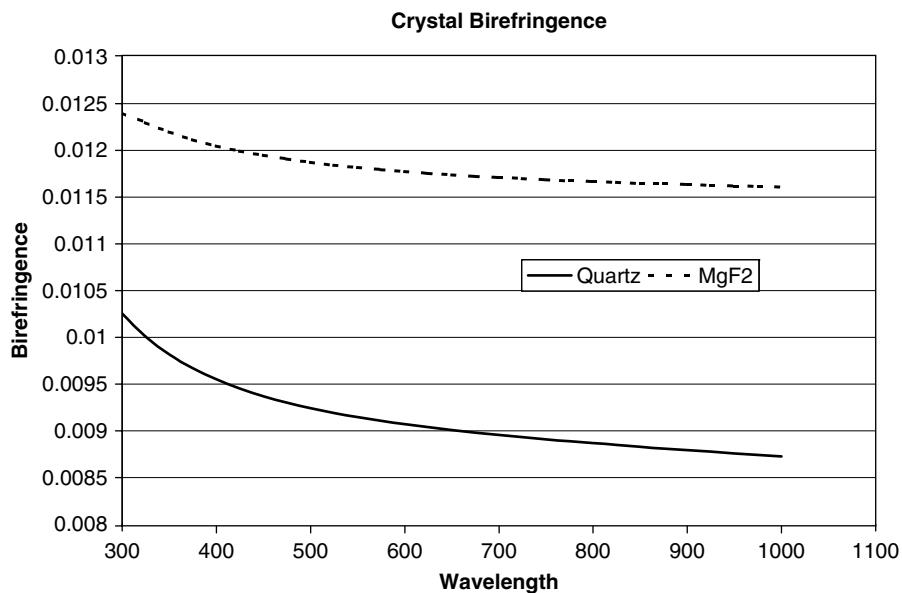
$$R = \beta d \quad (19.26)$$

where β = birefringence = $(n_e - n)$ and d is the thickness of the birefringent retarder. The result of this computation is the retardance in length units. The retardance in waves is retrieved by dividing this result by the wavelength. Since birefringence β is a function of wavelength, the retardance in length units will vary slowly with wavelength. However, the retardance in waves varies more quickly than the retardance in length units because of the division of the length retardance by wavelength. Figure 19.18 shows the variation of birefringence with wavelength for quartz and magnesium fluoride.

Retardance is also stated in radians or degrees of phase change. For example, a half-wave retarder has a retardance of 180° . An examination of the Mueller matrix for a retarder in Table 19.2 shows that the matrix is the same for a retardance of R waves and $n + R$ waves where n is any integer. This means that the effect of a half-wave or quarter wave

**Figure 19.18**

Wavelength Variation
of Birefringence of
Crystal Quartz and
Magnesium Fluoride



retarder on a polarized beam is the same as the effect of an $n + \frac{1}{4}$ quarter or $n + \frac{1}{2}$ half-wave retarder.

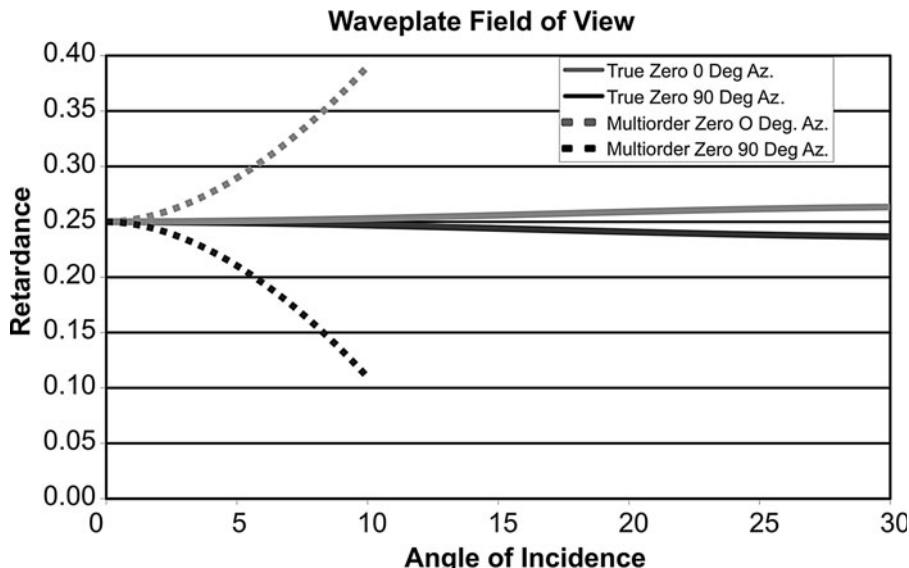
Multiwave versus Zero Order Retarders

We have just learned that the polarization change produced by a retarder is the same regardless of the number of integral waves of retardance and only the fractional remainder of waves matters. This is important because the birefringence of some commonly used waveplate materials is awkwardly high. For example, the birefringence of crystal quartz at a wavelength of 550 nm is 0.00917. The thickness of a quarter wave quartz waveplate for 550 nm is then only 15 μm . This is too thin and fragile to polish or handle easily. A 20.25 wave retarder has the same effect on the polarization as a quarter wave retarder and is 81 times as thick or about 1.21 mm thick. This is much sturdier. This thicker retarder is a multiwave retarder and the 15 μm one is a true zero-order retarder.

The mechanical advantage of the multiwave retarder is countered by some optical disadvantages. Figure 19.19 shows that the retardance is a much faster function of angle of incidence than for the true zero order retarder. In fact it is 81 times as sensitive to angle of incidence. This is

Figure 19.19

Retardance Dependence on Angle of Incidence for a True Zero-Order Waveplate and a Multi-order Waveplate. The Dependence for a Compound Zero Order Waveplate Is the Same as for the Multiorder One Shown Here if the Thicknesses are the Same



not always a disadvantage since a small tip of the multiwave retarder can adjust the retardance if needed.

Another disadvantage is that there is increased temperature variation of retardance in the multiwave retarder. The retardance decreases about $0.011\%/\text{ }^{\circ}\text{C}$ for quartz. If the retardance at $20\text{ }^{\circ}\text{C}$ is 20.2500 waves, then at $24\text{ }^{\circ}\text{C}$ it drops to 20.241 waves. This retarder will now perform the same as a 0.241 wave retarder. The retardance of the true zero order retarder only drops to 0.2499 waves for the same temperature increase. We have gained a factor of 81 in thickness for the multiwave retarder but have the same increase in temperature sensitivity. The multiwave retarder might still be the most cost effective choice if the waveplate is to be used in a collimated monochromatic beam in a stable laboratory environment.

The temperature insensitivity, but not the angular insensitivity, can be regained with a compound zero-order waveplate. This waveplate has two multiwave retarders with their fast axes crossed at 90° and their retardances differing by a quarter wave. The total retardance will be a quarter wave since the retardances of the two plates subtract. For example, one plate could have a retardance of 10 waves (0.5998 mm thick) and the second could have a retardance of 10.25 waves (0.6148 mm thick). The stackup thickness is the same as for the single 20.25 waveplate. The angular

sensitivity is the same as for the 20.25 multiorder waveplate. The damage threshold will be lower on the compound zero-order waveplate if the two component waveplates are joined with an adhesive.

A waveplate using a material with a lower birefringence can have both the low angular and thermal sensitivities of the true zero-order retarder and a thickness that makes it mechanically robust. There are no common crystal materials that have a significantly lower birefringence than quartz but some polymers can have a birefringence that can be adjusted to any value between zero and 0.03. These include polyvinyl alcohol, polystyrene, and polycarbonate. The birefringence is often adjusted by stretching the polymer to orient the long chain polymer molecules in a common direction. Under the right conditions the molecular orientation is maintained after stretching and the birefringence is stable with time. Often these polymer retarders are cemented between optically flat windows with an index-matching adhesive to achieve low transmitted wavefront distortion. The wavelength range of high transmission is usually less (325 to 2300 nm) than for the quartz retarders (180 to 2300 nm).

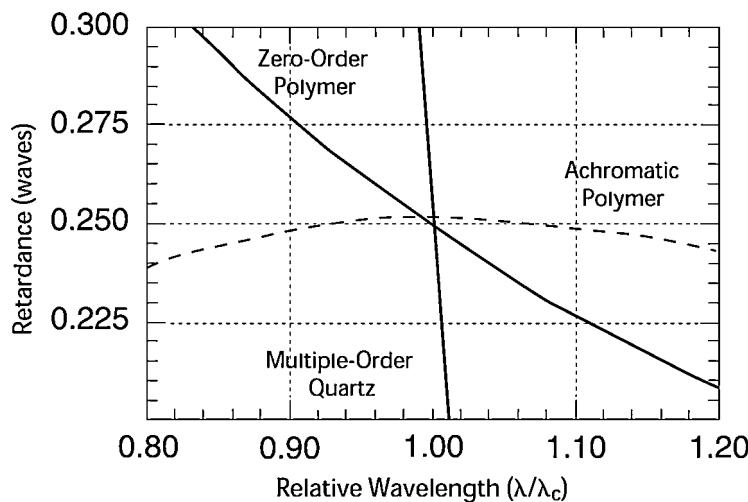
Achromatic Retarders

The retardance in waves of a retarder varies with wavelength mostly because of the division by wavelength of the retardance in length units. The wavelength variation of birefringence is a smaller secondary effect. For example, a zero-order half wave retarder for 550 nm can only be used over a wavelength range of 539 to 561 nm if the requirement is that the retardance remain within 0.01 waves of half-wave. Often this small range is insufficient and we must use an achromatic retarder. Figure 19.20 compares the wavelength dependence of one type of achromatic retarder to a zero order retarder.

One such retarder is the Fresnel rhomb. This is a prism device as shown in Fig. 19.21 that uses total internal reflection to produce a phase shift that is nearly constant with wavelength. Their drawback is that they are bulky, expensive and rather limited in clear aperture and angular acceptance. The beam is displaced in the quarter wave Fresnel rhomb but not in the half-wave rhomb. These devices are prone to retardance errors from stress birefringence if not mounted carefully, and the glass must be low strain because of the long path length. The retardance is a function only of the internal angle of incidence on the prism walls and

Figure 19.20

A Comparison of the Wavelength Dependence of Retardance for a Multiorder, Zero Order and An Achromatic Pancharatnam Quarter Wave Retarder



of the index of refraction of the glass. Figure 19.22 shows the wavelength dependence of retardance in a half-wave rhomb. This is the most achromatic type of retarder commonly available.

A second achromatic retarder is the bicrystalline achromatic retarder. These are usually compound zero-order retarders with each component waveplate made using a different crystal material. Usually the two crystals used are quartz and magnesium fluoride. The achromatic performance is a result of the interplay of the different wavelength dependences of

Figure 19.21

A Half-Wave Fresnel Rhomb Made of Two BK7 Glass Prisms. One of These Prisms can be Used as a Quarter Wave Retarder

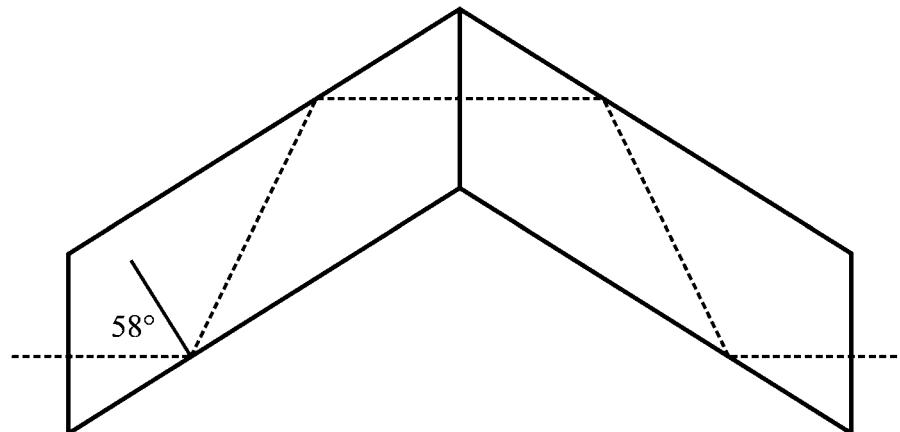
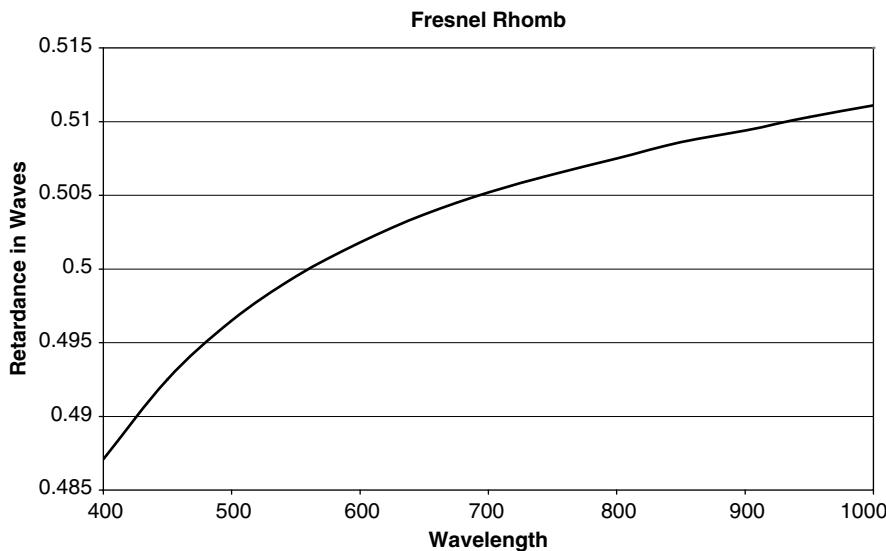


Figure 19.22

Wavelength Dependence of Retardance of a Half-Wave Fresnel Rhomb Made of BK7 Glass



birefringence of the two crystals in a manner somewhat analogous to an achromatic doublet lens using different glasses. The performance of one of these retarders is shown in Fig. 19.23.

A third achromatic retarder is the Pancharatnam retarder or generalizations of this design. The simplest Pancharatnam design uses three

Figure 19.23

Wavelength Dependence of Retardance of a Bicrystalline Achromatic Half-Wave Retarder Made of Quartz and Magnesium Fluoride

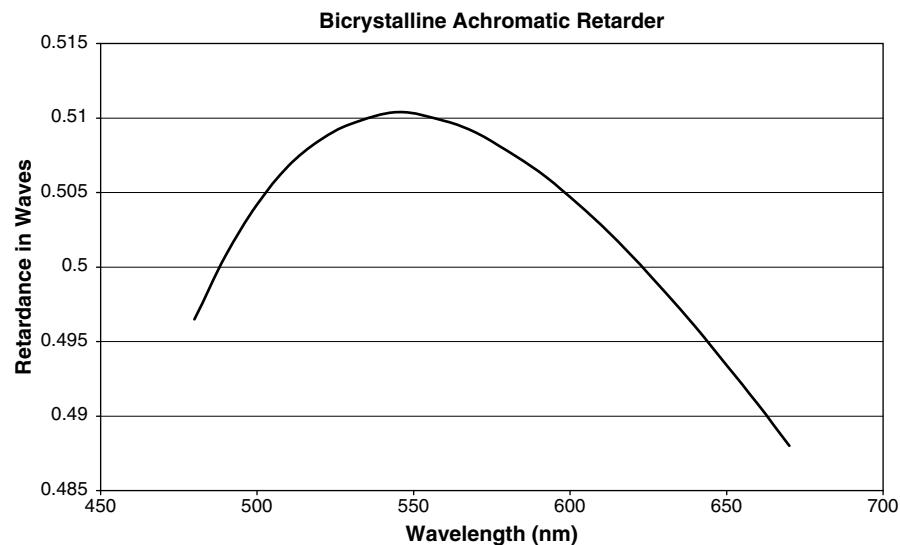
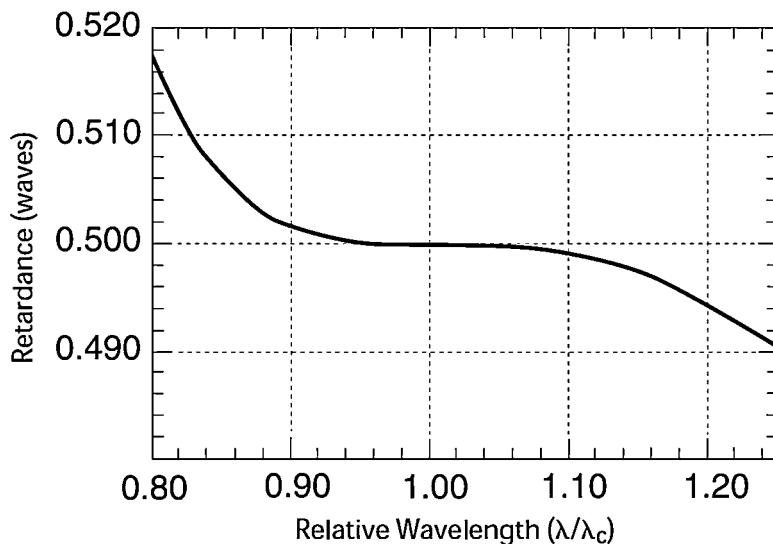


Figure 19.24

Wavelength Dependence of Retardance
for a Pancharatnam
Half-Wave Retarder



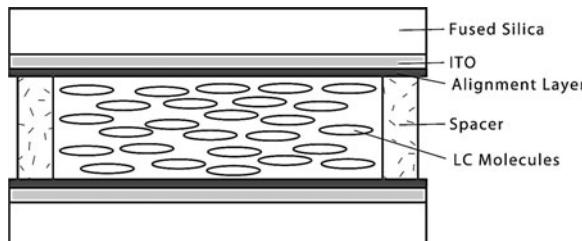
component waveplates of different retardances and fast axis orientations. The component waveplates are usually true zero-order polymers. These achromatic retarders have a lower damage threshold than the other types and there is a small variation in the fast axis direction of the retarder with wavelength. Figure 19.24 shows the wavelength dependence of retardance for a half-wave Pancharatnam retarder.

Variable Retarders

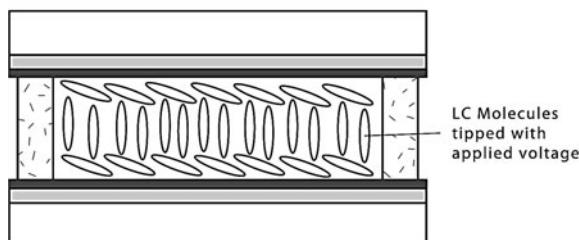
The most versatile retarders are electrically variable ones. These allow electrical selection of any retardance, within a range, for any wavelength for which the retarder transmits. The most common and popular ones use nematic liquid crystals as the birefringent material. These liquid crystals are uniaxial just as solid crystals are but the molecular and therefore the optic axis direction is electrically adjustable as shown in Fig. 19.25. The drive voltage is less than 20 V at about 1 to 2 kHz, square wave. Tilting the optic axis out of plane produces a change in the effective birefringence of the liquid crystal layer, which is only a few microns thick. The relationship between applied voltage and retardance is nonlinear as shown in Fig. 19.26. Response times range from about 3 to 50 ms depending on the retardance shift and the thickness of the liquid crystal layer. The retardance change with temperature is approximately $-0.2\%/\text{ }^{\circ}\text{C}$.

Figure 19.25

A Schematic Diagram of a Nematic Liquid Crystal Variable Retarder. The Optic Axis Direction is Parallel to the Long Axis of the Liquid Crystal Molecules. This Direction Changes with Voltage Applied to the ITO Transparent Conductive Electrodes



(a) Maximum Retardance ($V = 0$)

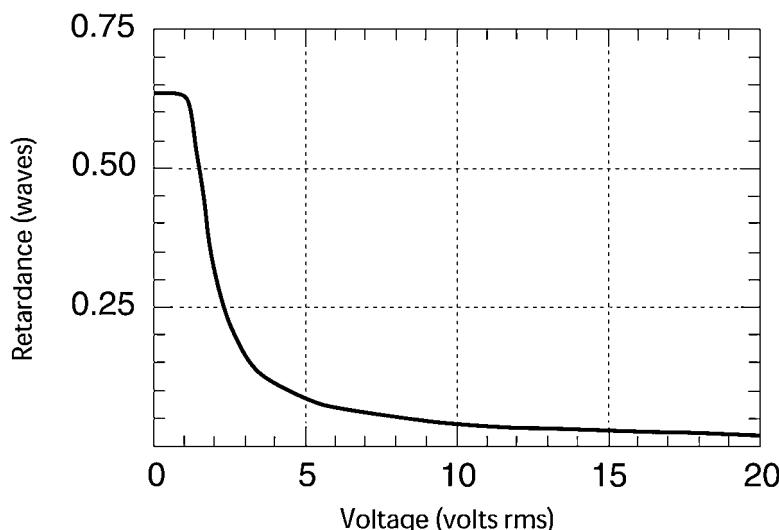


(b) Maximum Retardance ($V > 0$)

There are also liquid crystal retarders made using smectic A and smectic C liquid crystals that have a fixed retardance but have electrical control of the fast axis direction in the plane of the liquid crystal layer. These retarders can switch axis direction in less than 20 μs .

Figure 19.26

Retardation as a Function of Applied Voltage for a Liquid Crystal Variable Retarder. The Applied Voltage Is a 2 kHz Square Wave



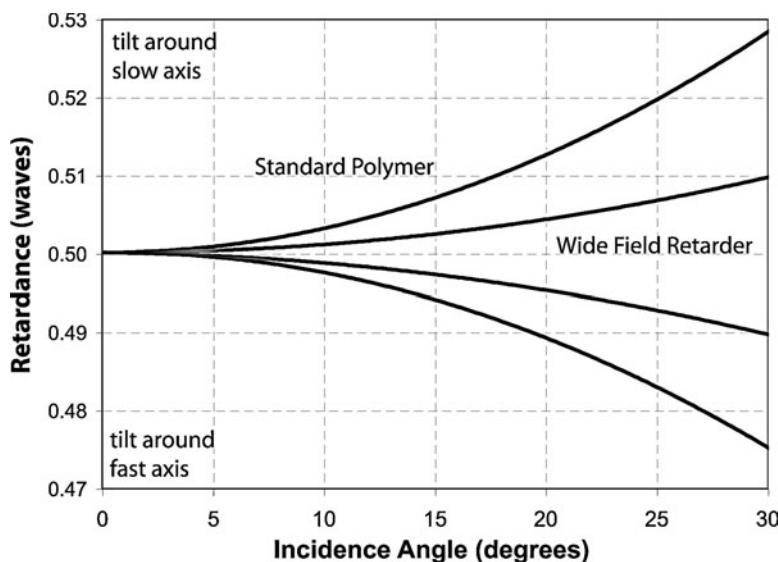
Pockels cells can be used as variable retarders but they require over 2000 V to reach quarter wave retardance at a wavelength of 633 nm and twice that to reach half-wave. The retardance is a linear function of applied voltage and the retardance is zero with no voltage applied. The active material is usually KD₂P which is potassium dideuterium phosphate crystal. The response times can be picoseconds but there can be ringing effects in the response if driven with a square wave voltage because the crystal is piezoelectric. These variable retarders cannot be held at a fixed retardance for more than about 100 ms without damaging the crystal and the electrodes. They must be run in a DC-balanced mode. When the sign of the voltage is reversed so is the sign of the retardance.

Special Retarders

For some applications the angular field of view of even the true zero-order waveplate is inadequate. The field can be further broadened as shown in Fig. 19.27 by using a combination of different birefringent materials in a manner beyond the scope of this chapter.

Sometimes there is a need for a retarder that has a lower temperature sensitivity. This can be met by using a combination of two materials in

Figure 19.27
Comparison of the Angular Change of Retardance for a Standard Polymer True Zero-Order Retarder and a Special Wide Field Retarder



a compound zero-order retarder configuration. The ratio of thicknesses of the two crystals is chosen to balance the temperature shifts in the two materials so that the net retardance stays fixed as the temperature changes. We have stated earlier that the retardance of a quartz retarder that is either compound or true zero-order drops by 0.01% for a temperature increase of 4°C. This can be decreased to less than 0.001%/°C with a two-material design.

Polarization Analysis of an Optical System

Given that many polarization phenomena may be present in an optical system, how is the engineer to determine what, if anything, needs to be done about it? What tools are available to help? What constitutes a polarization analysis of an optical system?

Sometimes the polarization characteristics of an optical system are specified by a customer or system engineer, in which case the polarization analysis consists of verifying that the optical design will meet the polarization specifications when built. Other times no polarization requirements are explicitly made, but it is expected that the optical designer will design a system in which the polarization behavior is consistent with the other optical specifications.

For example, the optical engine for a liquid crystal microdisplay projection system may have a high contrast requirement. There may be no explicit polarization requirement, but if the designer fails to understand the impact of the polarization behavior of the optical elements, then reaching high contrast may be impossible. Similarly, a space-borne earth-observing radiometer may have a radiometric uncertainty requirement. Since the light reflected from the earth can be highly polarized at the viewing angle of the sensor, the polarization dependence of the sensor responsivity may be important.

There is no one perfect method to analyze an optical system with regard to polarization behavior, but a general method that can prove useful is the following:

1. Examine the specifications or requirements of the optical system and note whether polarization behavior is either explicitly or implicitly important.

2. Determine whether the optical system contains highly polarizing optical elements (such as polarizers, beam splitters, and retarders) in order to function.
3. For each highly polarizing optical element examine the range of use in terms of wavelengths, angles of incidence, and any other parameters over which the polarization performance might vary. Consult the component vendor to verify the suitability of the component for the application.
4. Determine whether unpolarized or partially polarized light will be incident into the system, or whether any internal source is likely to be polarized.
5. Determine whether the responsivity of any detector in the system is highly dependent on the polarization state incident on it.
6. Identify portions of the optical train in which the polarization behavior is highly critical and portions in which it is not critical.
7. In the polarization critical regions look for the following:
 1. Mechanical and thermal environment
 2. Multilayer coatings
 3. High angles of incidence on lenses or mirrors
 4. High NAs
8. For each of these items analyze whether the polarization behavior is likely to cause a problem. This analysis may consist of a combination of examining vendor specifications, performing a polarization ray trace (see below), taking component or subsystem measurements, or making good engineering judgment based on experience.
9. If necessary, perform a polarization ray trace either with commercial software or through modeling using the Jones or Mueller calculus in order to verify that the system will meet its polarization requirements. Depending on the system and its requirements this may need to be very detailed or it may be very simple.

Polarization Ray Tracing

In some systems it is both desirable and possible to perform a full polarization ray trace. A polarization ray trace is an extension of conventional

ray tracing that computes the polarization state of the ray in addition to its direction and phase. It allows one to compute the power and polarization state of a ray at any optical surface in the system. In some systems a full polarization ray trace is desirable but not possible, due to the lack of information required to model each component. A full polarization ray trace traces rays through all elements of the optical system, from object to image or source to detector. Some systems require a polarization ray trace only of a subset of the optics.

Some commercially available lens design software incorporates polarization ray tracing to some degree. There are several levels at which ray tracing software can address polarization:

1. Dielectric films: Many software packages allow thin films to be placed onto glass or mirror surfaces, and they calculate the *s* and *p* transmission coefficients at each coated surface. They take into account the rotation of the *s* and *p* orientations with the plane of incidence on a ray by ray basis.
2. Birefringence: A small number of software packages allow propagation in uniaxial birefringent media. The user defines a glass type with a crystal axis direction and the ordinary and extraordinary refractive indices. The user must choose for a given ray trace whether to trace the ordinary ray or the extraordinary ray for each birefringent medium. Software that allows for biaxial or optically active media is rare.
3. Ideal polarization elements: Some software packages allow the insertion of ideal polarization elements, that is, perfect polarizers whose axis may be set by the user, or a perfect retarder, whose axis and retardance may be set by the user. Note that an ideal element may be easy to define at normal incidence, but not necessarily off-axis. For critical applications the off-axis behavior of the ideal model in software must be well-understood in order to compare it to the real element that will be used.
4. Nonideal arbitrary polarization elements: Software may allow the introduction of an arbitrary polarization element at a surface. The user enters, for example, the Jones matrix associated with the element. Given that the Jones matrix may be wavelength and angle dependent, it is often difficult to insert an appropriate polarization model into the software.

5. Real polarization elements: Models for real polarization elements, matching commercially available polarization components, almost never exist in commercial software. The polarization behavior is likely to be well-established only over a limited range of parameters (if at all), and one vendor's components may differ substantially from another's.
6. Ray splitting: At an isotropic to birefringent interface, the ray energy of an incident beam will be split into one reflected beam and two transmitted beams (the ordinary and extraordinary modes). At a birefringent to isotropic interface, the ray energy will split into two reflected beams and one transmitted beam. At a birefringent to birefringent interface the incident beam will split into four beams. A nonsequential ray trace engine can, in principle, manage the multiplication of rays and the polarization-dependent coupling coefficients.
7. High NA: As the assumptions of geometric optics begin to break down, the reliability of ray tracing software deteriorates at high NA. At least one major lens design program can be purchased with options that allow analysis at high NAs. Specialized software exists that performs diffraction calculations beyond the bounds of conventional geometric optics. These may be required for some high NA systems in which polarization behavior is critical.

POLARIZATION PITFALLS Polarization modeling in commercially available software has come a long way in the last decade, but its use still requires care. There are three significant pitfalls in using this software:

1. The local coordinate system representations for polarization are arbitrary. The user must understand the coordinate system assumptions that the software follows in order to interpret the polarization results of off-axis rays. Sometimes these assumptions are difficult to understand from the manuals.
2. Entering ideal components can give a useful “first order” approximation of the polarization behavior, but the detailed off-axis behavior of real polarization elements is rarely well-known. The results from the ideal model may be misleading.
3. Generally speaking, the polarization portions of commercially available software have been less thoroughly tested than other features. Calculations involving thin film stacks may be

considered very reliable. Other features may or may not have passed the tests of time.

In conclusion, commercially available optical analysis software can be a very valuable tool in the polarization analysis of an optical system, but it is still somewhat limited in applicability. The results of a polarization ray trace must be interpreted with care. Polarization components are never ideal, and thermal environments may be unknown or dynamic.

Minimizing Polarization Problems in Optical Design

In the design of a polarization critical optical system several rules of thumb or design tricks can be applied, although every system is of course different:

1. Analyze your requirements and identify polarization critical regions of the optical system. Keep these regions as small and simple as possible.
2. In polarization critical regions keep the angles of incidence on surfaces low, if possible.
3. In areas subject to thermal gradients pay close attention to the method of lens mounting in order to minimize stress birefringence. Difficult locations may require use of a glass type that is resistant to stress birefringence.
4. Watch out for plastic lenses, which may have significant built-in stress birefringence if injection molded.
5. If a folded optical system is required, it may be helpful to fold twice, such that the *s* orientation at the first mirror aligns, on reflection, with the *p* polarization of the second mirror. If the mirrors have the same material and coating, then their polarization characteristics will compensate one another for the on-axis ray. For off-axis rays the cancellation will not be perfect, but may be good enough for the application.
6. Introducing a depolarizer will make the polarization dependence of the subsequent elements irrelevant.
7. Where possible, obtain detailed polarization specifications from component vendors. Where impossible, measure what is necessary.

Polarization as a Tool in Optical System Design

Polarization can be both a problem and a solution in optical system design. Polarizing components can be a solution for controlling:

1. System transmission
2. Optical beam intensity
3. Beam direction
4. Spatial distribution of light
5. Phase and phase distribution in a light beam
6. Filter transmission wavelength, and
7. Mechanical motion

We will give examples of all these control functions. Interestingly, the use of electrically variable polarization devices, usually variable retarders, enables nonmechanical polarization control for all seven of these functions. Thus, polarization provides a pathway for nonmechanical, that is, electrical, control of these optical functions.

Controlling System Transmission and Optical Beam Intensity

Rotation of a pair of linear polarizers is a simple obvious mechanical way to control transmission. This follows the law of Malus as given in Eq. 19.21 and with typical dichroic sheet polarizers permits transmission variable attenuation at visible wavelengths over a range of three or more orders of magnitude. Polarizer rotation for attenuation can cause undesired system response because of either source polarization or detector polarization sensitivity.

A half-wave retarder rotated between fixed polarizers avoids these problems. This has the disadvantage that the attenuation will be strongly wavelength dependent unless the half-wave retarder is achromatic. An electro-optic solution is to replace the half-wave retarder with an electrically variable retarder. This removes the need for mechanical motion but is another wavelength dependent solution. A fourth solution uses a twisted nematic liquid crystal cell between parallel polarizers. This

solution is more achromatic than using a half-wave retarder but is still limited in its wavelength range. Clearly, each solution has different benefits and drawbacks and the best solution will depend on the system requirements.

Controlling Beam Direction

Figure 19.28 also shows one example of the use of polarization to control beam direction with a polarizing beamsplitter cube. The beam direction is binary in this example and the choices of mechanism for control of polarization direction and thus beam direction are the same as those listed for controlling beam intensity. Although there are only two selections for beam direction, the angle between the directions can subsequently be varied by prisms or mirrors. The combination of an electrically variable retarder and a polarizing beamsplitter as shown in Fig. 19.29 can also act as an electrically variable beamsplitter. The splitting ratio is widely variable from 1000:1 (transmitted to reflected) to 1:100 with commonly available optical components and with monochromatic light.

Figure 19.28
Beam Direction Control by Rotating a Half-Wave Retarder Following a Fixed Polarizer. This Configuration is Also a Mechanically Variable Beamsplitter

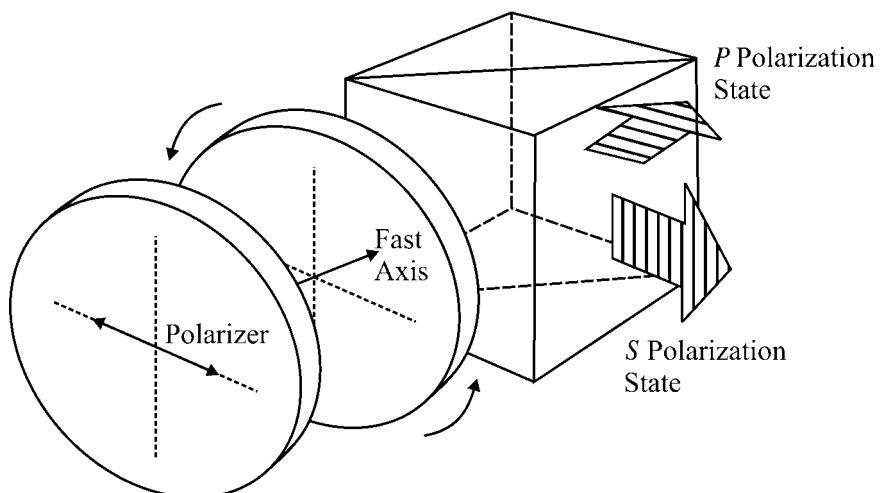
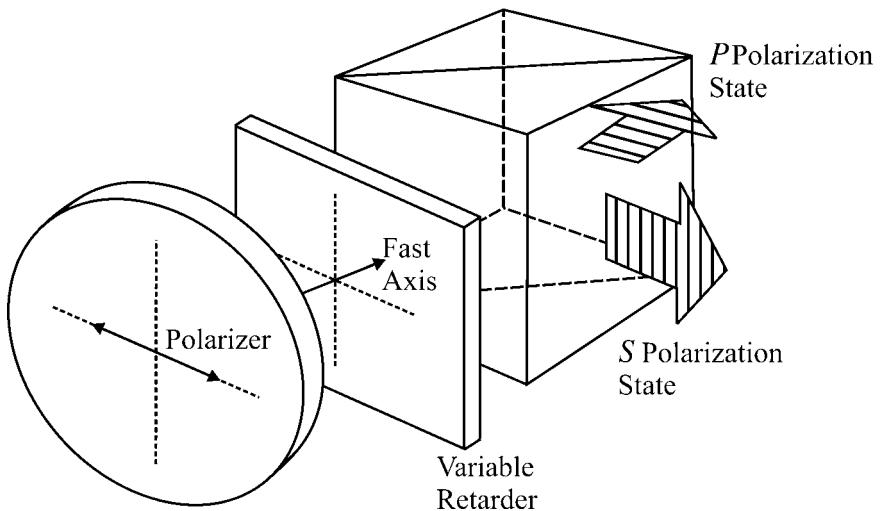


Figure 19.29

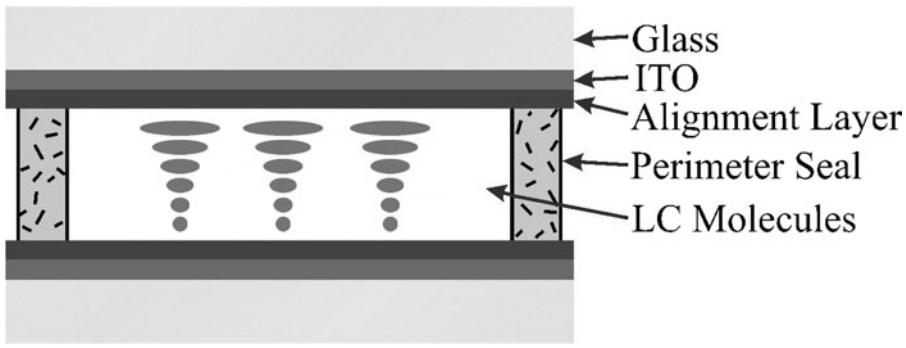
An Example of Controlling Beam Direction by Electrically Controlling Polarization. This Also Acts as An Electrically Variable Beamsplitter



Controlling the Spatial Distribution of Light

By far the most common example of this application of polarization is the LCD. In this case there are thousands or even a few million individually electrically controlled polarization elements such as on a computer screen or in a video projector. The individual picture elements or pixels can be liquid crystal variable retarders or more commonly they are twisted nematic or supertwisted nematic liquid crystal elements between polarizers. These latter two pixel types operate on the concept of adiabatic following of polarization through the twisted uniaxial crystal structure. The linear polarization direction follows the crystal optic axis as it moves through the liquid crystal layer as shown in Fig. 19.30. The rotation is approximately 90° in the twisted nematic pixel and approximately 270° in the supertwisted nematic pixel. Applying a voltage to the cell removes the twist and therefore removes the adiabatic following of the polarization direction. There are other liquid crystal configurations that are used in displays, depending on important parameters that must be met such as angular contrast variations, response time, and degree of achromaticity desired.

There are other systems besides displays that require the producing a spatially varying distribution of light and usually this distribution

**Figure 19.30**

A Schematic Diagram of a Twisted Nematic Liquid Crystal Cell. The Optic Axis Direction Is in the Long Direction of the Liquid Crystal Molecules and This Direction Rotates by 90° between the Top and Bottom Cell Walls. A Voltage Applied to the ITO Electrodes will Remove the Twist and Change by 90° the Direction of Polarization of Light Emerging from the Cell

must be under electrical control. Examples include optical correlators, pulse shapers for femtosecond lasers and Hadamard spectroscopy systems. Liquid crystal devices are used in these applications as well.

Controlling Phase

The optical phase delay in a linear retarder is dependent on the azimuth of linear polarization of the incident beam as we have discussed. Light polarized along the fast axis is delayed less than light polarized along the slow axis. Therefore changing the input polarization changes the phase delay. We can also adjust phase by adjusting the effective birefringence of an electrically variable retarder since the optical path distance through the retarder is just the product of the physical thickness and the index of refraction. For example, in a nematic liquid crystal variable retarder the birefringence is varied electrically by varying the index of refraction along the slow axis of the liquid crystal. Electrically induced phase delay changes in nematic liquid crystals can be as small as a nanometer and as large as about $10 \mu\text{m}$.

Often the system requirement is for a spatial and time variable phase and liquid crystals can provide this, just as in displays. Examples of these systems include beam steerers, laser pulse shapers, and active and adaptive optical imaging systems such as those used on astronomical telescopes.

Controlling Transmission Wavelength

The retardation, in waves or fractional waves of light, of a retarder is a strong function of the wavelength of light for most types of retarders. The wavelength dependence limits the wavelength range over which the variable attenuators discussed in this section achieve a desirable performance. In other words, these attenuators work well at some wavelengths and not at all at others. This wavelength dependence becomes a benefit rather than a hindrance when wavelength selective transmission is a desired system function. For a multiwave retarder between parallel polarizers transmission the transmission T is

$$T = \cos^2(R/2) \quad (19.26)$$

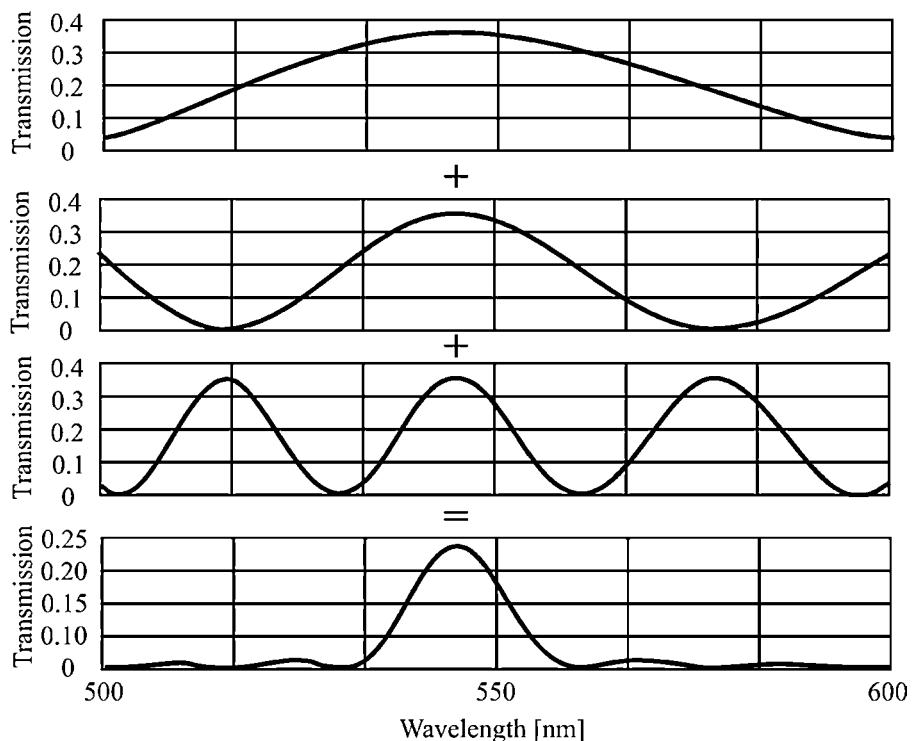
where R is the the retardance in angular units. Peaks occur at wavelengths where the retardation is an integral number of waves and minima occur where the retardation is an integer plus a half-wave.

By itself the optical transmission from the 4 waves or 2180 nm retarder shown in Fig. 19.31 is not very interesting since there are adjacent transmission peaks in the blue at 436 nm in the red at 727 nm and the peak at 545 nm is quite broad. However, by adding more *stages*, that is multiwave retarders between polarizers, the transmission peak becomes narrower as shown in the bottom curve in Fig. 19.31. This type of polarization interference filter is called a *Lyot filter* after its inventor, Bernard Lyot. In the example of Fig. 19.31 we have added Lyot filter stages with retardations of 4360 and 8720 nm or 8 waves and 16 waves, respectively, at a wavelength of 545 nm. This combination of three Lyot filter stages gives a transmission bandpass full width at half maximum of approximately 14 nm. Adding a 2-wave stage will remove the adjacent transmission peaks at 436 and 727 nm. Adding a 32 wave stage will narrow the bandpass width to 7 nm. Polarization interference filters of the Lyot type have been built with bandpass widths as narrow as 0.0125 nm.

There are several other types of polarization interference filters that are more difficult to explain that have the advantage of reducing the number of polarizers required and therefore producing higher peak transmission. Two examples are the Solc filter and the Evans split element filter. The bandpass wavelength in the Lyot filter is electrically adjustable if the retarders are electrically variable. The bandpass wavelength is mechanically adjustable if the retarders are fixed by rotating polarizers or a quarter wave plate.

Figure 19.31

Transmission versus Wavelength for a 4 Wave (Top), 8 Wave and 16 Wave Retarder, Each between Parallel Polarizers, at a Wavelength of 545 nm. The Combination of These Retarders and Polarizers is a Band-pass Filter (Bottom)



Controlling Mechanical Motion

In the past decade or so new methods have developed to use polarization to control motion. Contraction direction in some polymers is dependent on the polarization as shown in Fig. 19.32. Response time is less than 10 seconds. This could develop into applications for artificial muscles for miniature robots, for example. Optical tweezers benefit from polarization for the manipulation of microscopic objects such as cells. Bose-Einstein condensates are formed by laser cooling of atoms in systems where polarization control is important.

Summary

In most optical systems polarization is not an important factor, but in those systems in which it is important it can be critical. An imaging system with wavefront error of $\lambda/10$ is usually considered diffraction limited

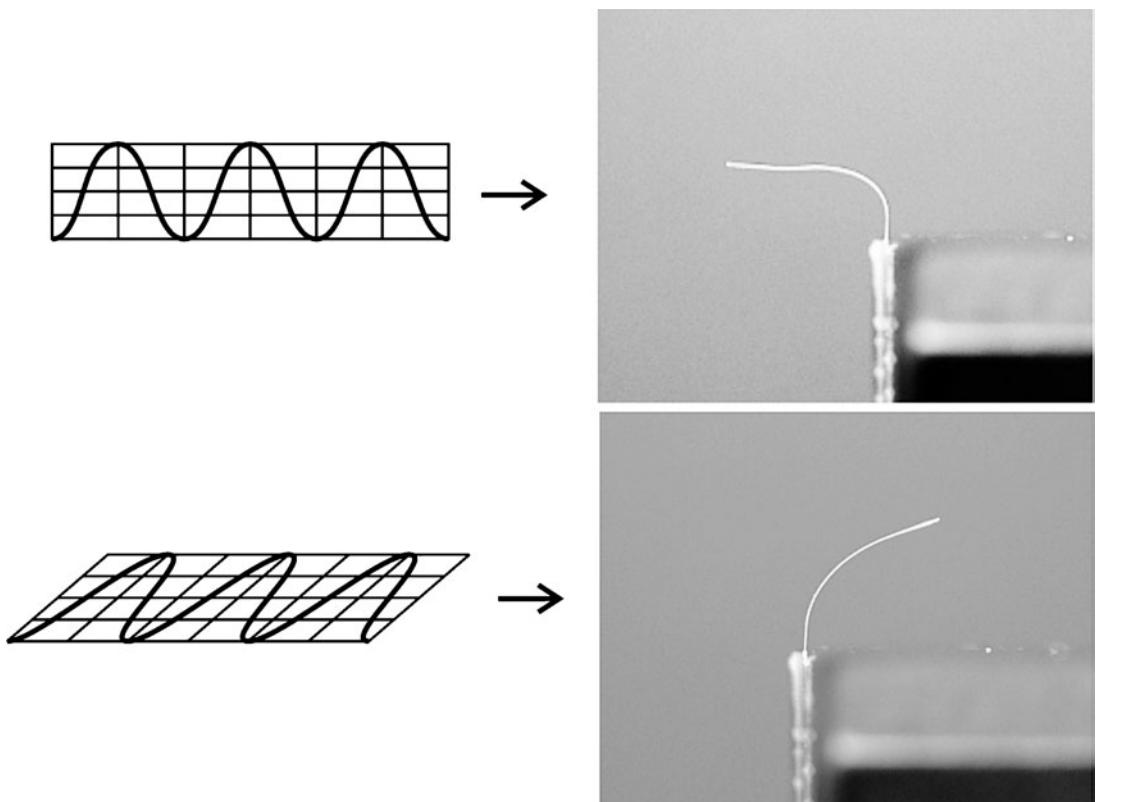


Figure 19.32

A Photograph of a Polymer Strip with the Direction of Curl Controlled by the Direction of Linear Polarization of the Incident Light (*Courtesy of BEAM Engineering, Inc.*)

and adequate for all but the most demanding applications. But stray retardance of $\lambda/10$ in a polarization critical region of an LCD projection system can reduce contrast from 1000:1 to 10:1, rendering it unusable.

With some knowledge of the physical and mathematical principles of polarization, most optical designers can perform a polarization analysis of an optical system, perhaps with the assistance of commercial software. However, there are some systems whose polarization complexity or whose extremely demanding requirements will be difficult for the non-specialist to handle. Whether to apply the Jones calculus, the Mueller calculus, polarization aberration theory, or some mixture of methods to a polarization problem is not possible to state generally. Knowledge of

the extensive idiosyncrasies of various polarization components and how to deal with them is also beyond the scope of this chapter, as is the entire subject of precision polarimetry and the interpretation of polarimetric data.

The difficulties of polarization for the optical designer are of two types. First, the polarization mathematics applied to optical systems containing off-axis rays (that is, nearly all optical systems) is rather cumbersome, involving the complexities of local coordinate system representations and the difficulty of expressing three dimensional vectors in two-dimensional projections. Software exists that reduces but does not eliminate the computational difficulty. Second, the polarization behavior of real optical elements is often very complex and not necessarily known even by the vendor. This often necessitates polarimetric measurements on critical components, which in some cases can be a more difficult and time-consuming task than the optical design and analysis of the entire optical system.

Bibliography

- Azzam, R.M.A., "Ellipsometry," in: *Optical Society of America Handbook of Optics*, Bass, M., Ed., Vol. 2, Chap. 27, McGraw-Hill, Inc., NY, 1997.
- Azzam, R.M.A., and Bashara, N.M., *Ellipsometry and Polarized Light*, North-Holland Publishing Company, Amsterdam, 1977.
- Bennett, J. M., "Polarization," in: *Optical Society of America Handbook of Optics*, Bass, M., Ed., Vol. 1, Chap. 5, McGraw-Hill, Inc., NY, 1995.
- Bennett, J. M., "Polarizers," in: *Optical Society of America Handbook of Optics*, Bass, M., Ed., Vol. 2, Chap. 3, McGraw-Hill, Inc., NY, 1997.
- Collett, E., *Field Guide to Polarization*, SPIE Vol. FG05, SPIE Press, Bellingham, Wash., 2005.
- Chipman, R. A., "Polarimetry," in: *Optical Society of America Handbook of Optics*, Bass, M., Ed., Vol. 2, Chap. 22, McGraw-Hill, Inc., NY, 1997.
- Goldstein, D., *Polarized Light*, 2d ed., Marcel Dekker, Inc., New York, 2003.
- Kliger, D.S., Lewis, J.W., and Randall, C.E., *Polarized Light in Optics and Spectroscopy*, Academic Press, Inc., San Diego, 1990.
- Shurcliff, W.A., *Polarized Light*, Harvard University Press, Cambridge, MA, 1962.

This page intentionally left blank

Optical Thin Films

Introduction

Optical thin films have become an integral part of almost all optical components and systems manufactured today. Their primary function is to govern the spectral composition and the intensity of the light transmitted or reflected by the optical system. Properly applied to various optical surfaces in a given system, optical coatings can greatly enhance image quality and provide for a convenient way of spectrally manipulating light.

Since light behaves according to the laws of electromagnetic waves, the interaction of light with the media that it travels through, or is reflected from, is directly related to its wave nature, primarily the phenomena of *interference* and *polarization*. Whenever light interacts with a structure of thin films, interference occurs, and a degree of polarization will be a function of the angle of incidence. At normal incidence, no polarization will take place, unless the light is transmitted through a birefringent material (polarizing material, like some crystals and plastics). Besides polarization, at an oblique incidence there is a *spectral shift* of the reflectance or transmittance characteristic toward the shorter wavelength. This is due to the optical path difference between the waves reflected from either side of the film structure. This optical path difference is directly proportional to the cosine of the angle of refraction through the coating.

For an optical designer, besides the fact that the interference and polarization are the most fundamental physical principles in the theory

of thin films, an important characteristic is the amount of energy loss, or the *light absorbed* in the coating. In general, for any coating there is a relationship between the transmittance, T , the reflectance, R , and the absorptance, A , in the form of

$$T + R + A = 1 \quad (20.1)$$

where $0 \leq T, R, A \leq 1$.

For materials that are commonly known as *dielectrics* the coefficient A in Eq. (20.1) is very close to zero, and they basically do not absorb any light. On the other hand, *metals*, besides being highly reflective (90 to 98%), act as light attenuators, and their coefficient of absorption is always greater than zero.

We will refer to Eq. (20.1) later on when we discuss different categories of optical thin films.

Designing Optical Coatings

Without getting into deep analysis of design methods of optical thin films, let us point out that the main building blocks in designing optical coatings are *quarter-wave optical thickness* (QWOT) layers of different materials. The high, medium, and low refractive index QWOT materials are usually denoted as H , M , and L , respectively. If there are two QWOT layers of the same material next to each other, they form a *half-wave optical thickness* (HWOT) layer. If only a fraction of QWOT appears in a design, say one half of H , it is represented as either $0.5H$ or $H/2$.

The long expressions for some designs can be represented in concise form. For example, a 15-layer longwave-pass filter on BK7 glass given by

$$\text{BK7} \left| \frac{H}{2} L H L H L H L H L H L \frac{H}{2} \right| \text{air}$$

can be written as

$$\text{BK7} \left| \left(\frac{H}{2} L \frac{H}{2} \right)^7 \right| \text{air}$$

where H and L refer to high and low index materials, such as TiO_2 and SiO_2 .

In principle, the computer programs that assist thin-film engineers in designing optical coatings are very similar to those used by optical designers. Optical design programs are more complex because there are

more variables (such as thickness, radius of curvature, refractive index) to simultaneously vary during the optimization. Further, they have a wider spectrum of the target functions to be satisfied at the end of the optimization (either in the form of the aberration functions, wavefront distortion, optical path difference, or the minimum spot size). Thin-film programs, on the other hand, deal with fewer variables (very often only thickness, rarely refractive index), and their target functions are usually in the form of either reflected or transmitted light intensity.

Thin-film computer programs are essential mathematical tools that enable coating engineers to efficiently, and in some cases very quickly, arrive at the best and most economical design once the problem has been formulated. But to successfully apply this math tool to coatings that are manufactured with high reproducibility, it is the engineer's knowledge of the coating materials and processes that determines the coating's final quality and conformity to the spectral and environmental requirements.

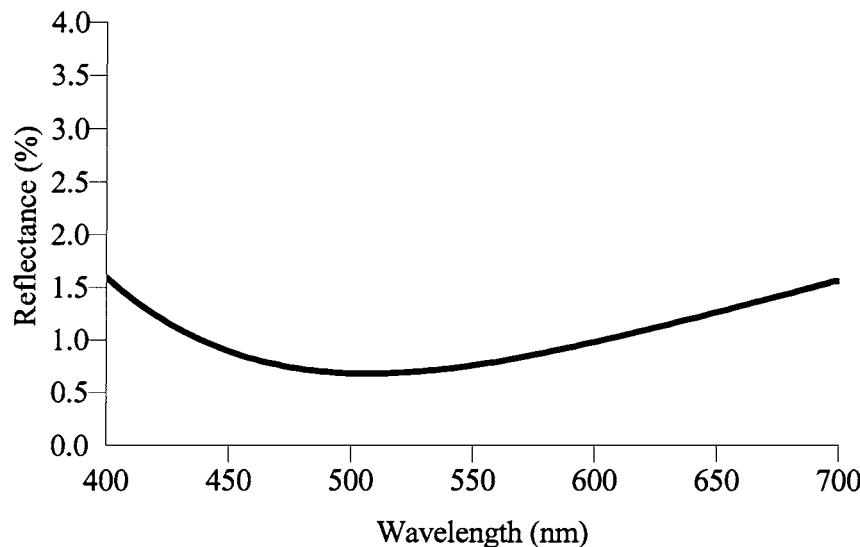
Various Categories of Optical Coatings

The most widely applied optical coating is the *antireflection (AR)* coating. Its primary purpose is to reduce the amount of reflected light from the optical surface. Its secondary role is to enhance physical and chemical properties of the surface to which it is applied.

Typically, uncoated glass has between 4 and 8% reflection from the surface. This can be reduced to about 1.0% reflection in the visible by applying a single layer of QWOT low-index material, usually magnesium fluoride (Fig. 20.1). A three-layer design can reduce the reflection in the visible even further (Fig. 20.2). The first layer consists of a QWOT medium-index material (for example, Al_2O_3) next to the glass. The second layer is a HWOT high-index material (for example, Ta_2O_5). The third layer is a QWOT low-index material (for example, MgF_2) as a top layer next to the air. This three-layer design falls in the category of the *broadband (BB)* antireflection coating, often denoted as BBAR coating.

If only one wavelength is considered, a two-layer design of high- and low-index materials will bring the reflection down to virtually zero value. With the layer next to the glass fairly thin (high-index material) and the layer facing the air side (low-index material) somewhat greater

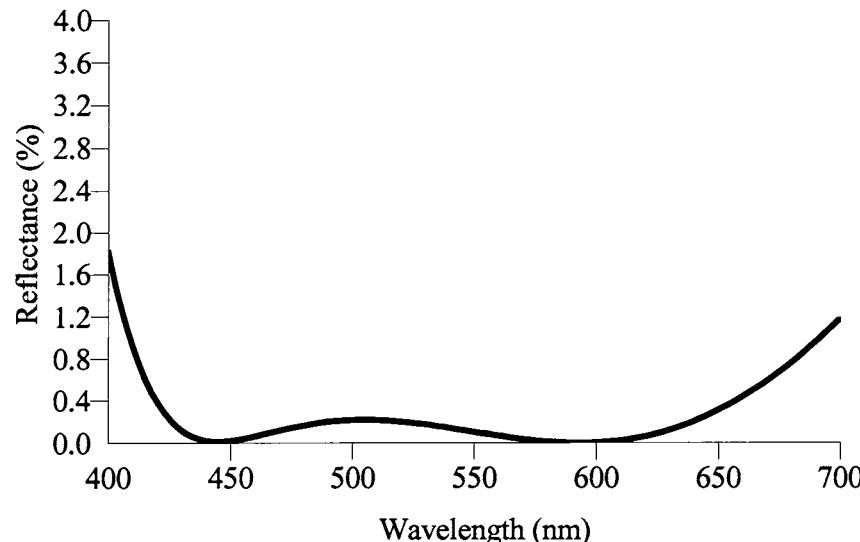
Figure 20.1
Computed
Reflectance at Normal Incidence of a Single Surface of SSK4 Glass ($n = 1.62$) Coated with a Single Layer of Magnesium Fluoride ($n = 1.38$) of Optical Thickness One Quarter-Wave at 510 nm. Design: SSK4| L |Air, AOI = 0°



than a QWOT, a relatively broad minimum can be obtained (Fig. 20.3). These coatings are usually called *V* coatings.

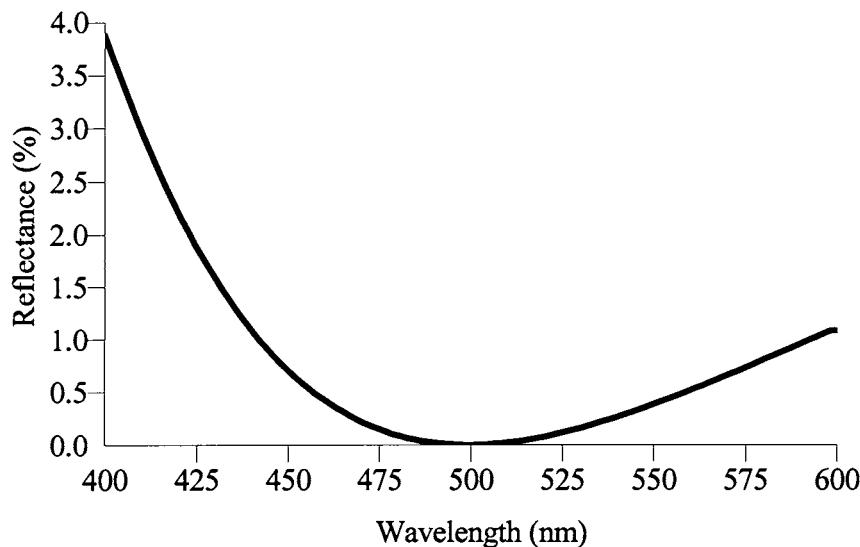
For much broader antireflective coverage that would include the visible and a near-infrared region, many layers of high- and low-index materials are required. Their thicknesses are computer optimized and monitored

Figure 20.2
Three-Layer Antireflection Coating on BK7 Glass ($n = 1.52$). (Design: BK7| MHHL |Air at 510 nm, AOI = 0°, $n_H = 2.126$, $n_M = 1.629$, $n_L = 1.384$)

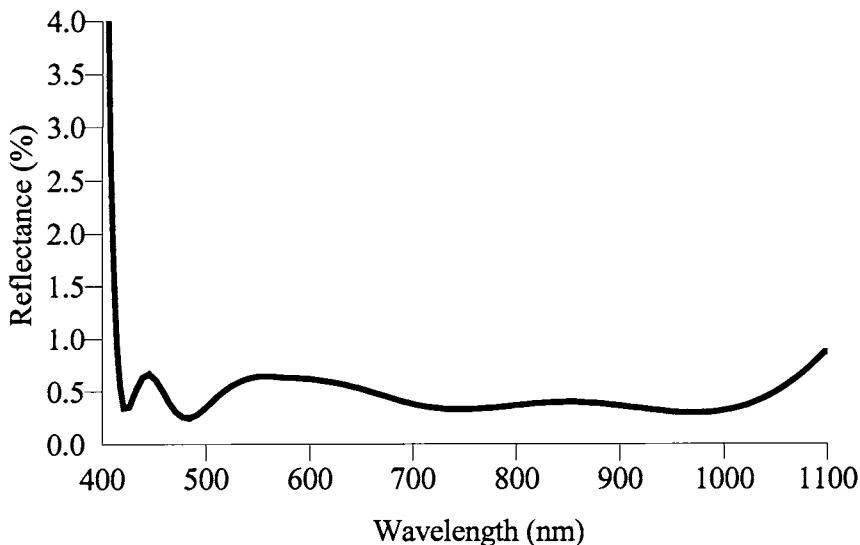


**Figure 20.3**

The Reflectance of a Two-Layer Antireflection Coating on BK7 Glass. (Design: BK7|0.2681H 1.2702L|Air at 500 nm, AOI = 0°, $n_H = 2.127$, $n_L = 1.384$)

**Figure 20.4**

Eight-Layer Antireflection Coating on BK7 Glass ($n = 1.52$). The Coating Consists of Two Materials of High- and Low-Refractive Index. This Design Has Been Computer Optimized and Has a Few Thin Layers (~20 nm) That Can Only Be Quartz Monitored. The Angle of Incidence Is 0°



Another class of widely used thin-film coatings is the *metallic mirror*, usually consisting of aluminum. Aluminum is a relatively soft metal, so the coating is often protected with silicon dioxide. The reflectance of this coating is about 90%, but can be further increased by adding a few more layer pairs of high- and low-index materials (for example, TiO_2 and SiO_2) to boost reflectance to about 97 to 98% (Fig. 20.5). Since aluminum is a metal, there is a slight light loss associated with its use. This light loss, or absorption, is manifested as heat released within the coating. In certain applications, such as high-power lasers, mirrors should be free of absorption to a very high degree. This is achieved through the use of all-dielectric mirrors.

Dielectric mirrors consist of the sequence of the alternating high- and low-index materials (for example, TiO_2 and SiO_2). The more layer pairs in the stack, the higher the reflectance. *Cold mirrors* reflect shorter wavelengths and transmit longer wavelengths (Fig. 20.6). *Hot mirrors* transmit shorter wavelengths and reflect longer wavelengths (Fig. 20.7).

As in the field of electronic circuits, there are many different interference filters in a variety of optical applications. Sometimes the goal is to separate one portion of the spectrum from the other. This separation can be done at either normal incidence or oblique incidence. Whatever the case, the solution will be in the form of an *edge filter* or some kind of *dichroic beamsplitter*.

Figure 20.5
The Reflectance of an Enhanced Aluminum Mirror with Four Layers of High- and Low- Index on Top of Aluminum. The Thickness of the Aluminum Layer Is 80 nm. (Design: Al | 0.8LHLH | Air at 520 nm, AOI = 0°, $n_H = 2.446$, $n_L = 1.459$)

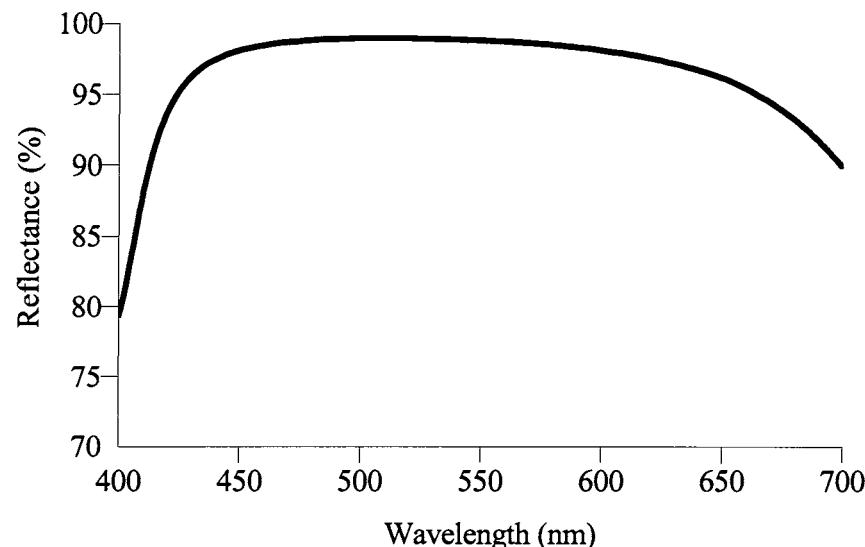
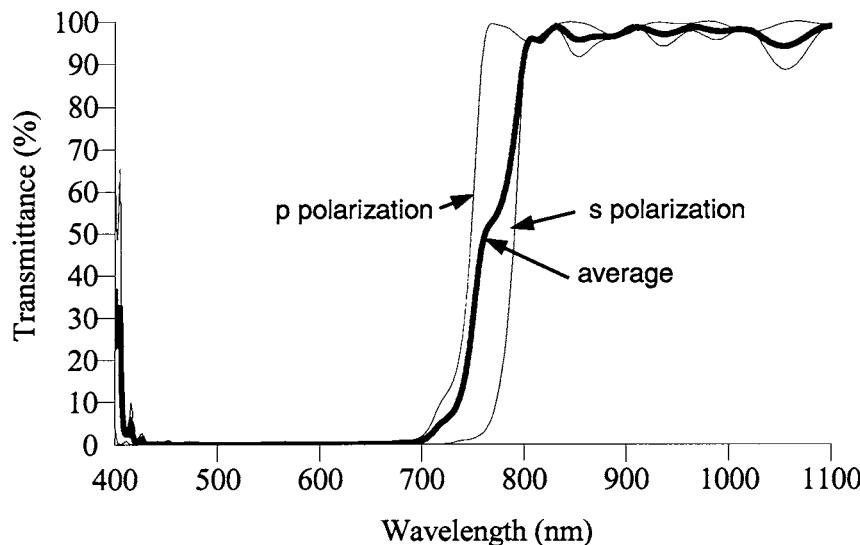


Figure 20.6
 The Transmittance of 52-Layer Cold Mirror at a 45° Angle of Incidence. Design Is Given in Phase Thicknesses (Degrees), and P_H and P_L Refer to TiO_2 and SiO_2 , Respectively.
 [Design:
 $\text{Air}[(100^\circ P_L)(74^\circ P_L)$
 $74^\circ P_H]^8[90^\circ P_L 90^\circ P_H]^8$
 $(108^\circ P_L 108^\circ P_H)^7$
 $(105^\circ P_L)(102^\circ P_H)(98^\circ P_L)$
 $[90^\circ P_H][98^\circ P_L](23^\circ P_H)$
]BK7 at 538 nm]



When there is a need to pass just one narrow bandwidth and reflect a portion of the spectrum on either side of it, use should be made of a *narrowband* interference filter, often called the *Fabry-Perot filter* (Fig. 20.8).

Recently, another class of interference filters has become of great importance in laser and fiber-optic applications: *notch filters*. They reflect

Figure 20.7
 Calculated Transmittance of a 44-Layer Computer Optimized Hot Mirror. [Design:
 $\text{Air}[(1.07L)(2H2L)]^8$
 $(2.6H)(2.64L)(2.8H)$
 $(2.46L)(2.14H)(2.2L)$
 $(2.6H)(2.6L)^3(2.6H)$
 $(2.74L)(2.9H)(2.9L)^5$
 $(2.74H)(3.08L)(0.4H)$
]BK7 at 415 nm,
 AOI = 0°, $n_H = 2.239$, $n_L = 1.463$]

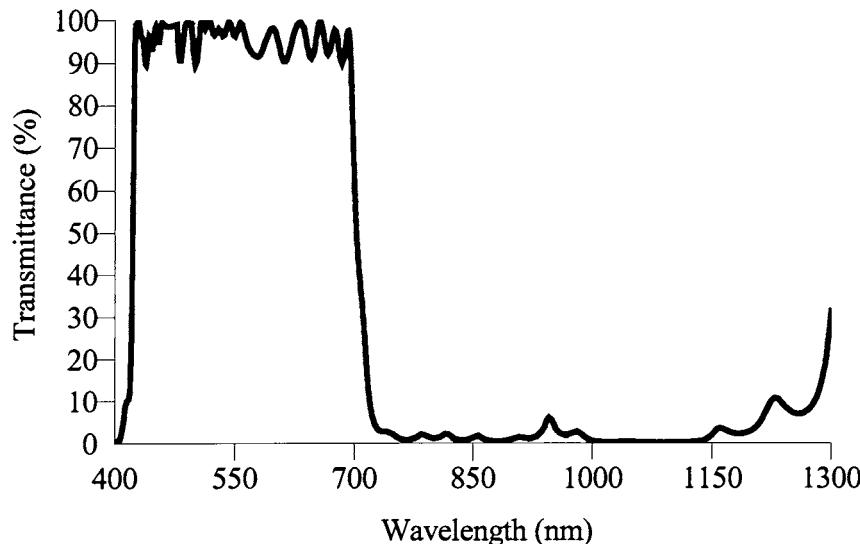
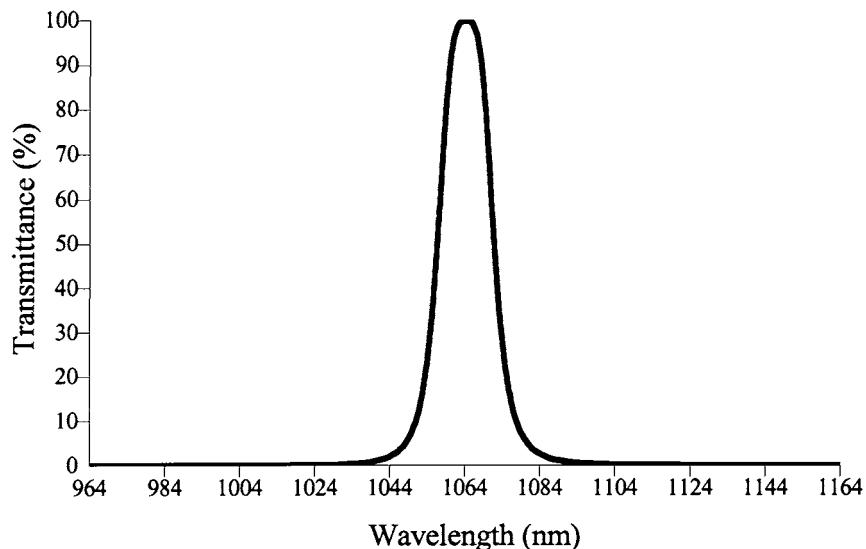
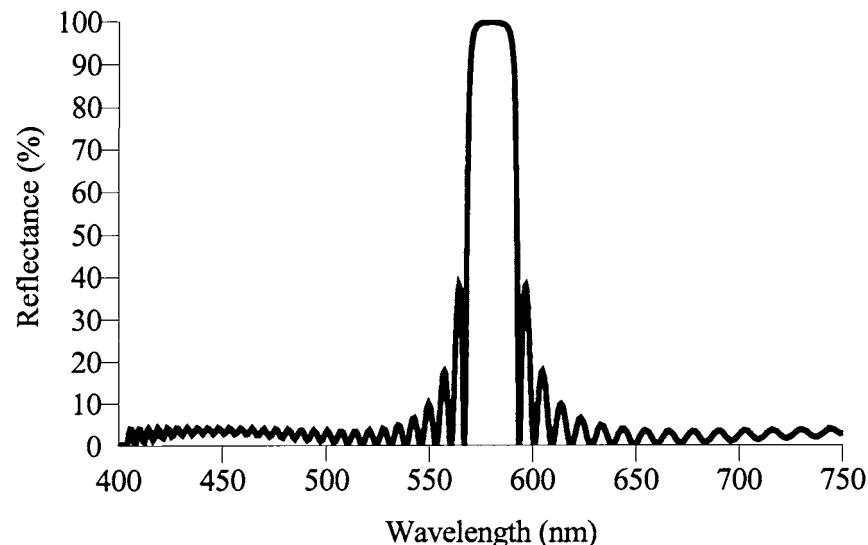


Figure 20.8

The Double-Cavity Fabry-Perot Narrowband Interference Filter at 0° Angle of Incidence [Design: BK7|(HL)³HH(LH)³L |(HL)³HH(LH)³|BK7 at 1064 nm, AOI = 0°, $n_H = 2.253$, $n_L = 1.449$]

**Figure 20.9**

Computed Reflectance of a Single-Notch Filter [Design: BK7|(L3M)³¹4L|Air at 580 nm, AOI = 0°, $n_M = 1.626$, $n_L = 1.457$]



With the advent of new polarizing devices in the area of electronic imaging, *polarizing beamsplitters* have become of significant importance. Their role is to maximize the s and minimize the p-reflectance of the unpolarized (randomly polarized) light over the narrow or broadband spectral region. The degree of polarization in transmission is

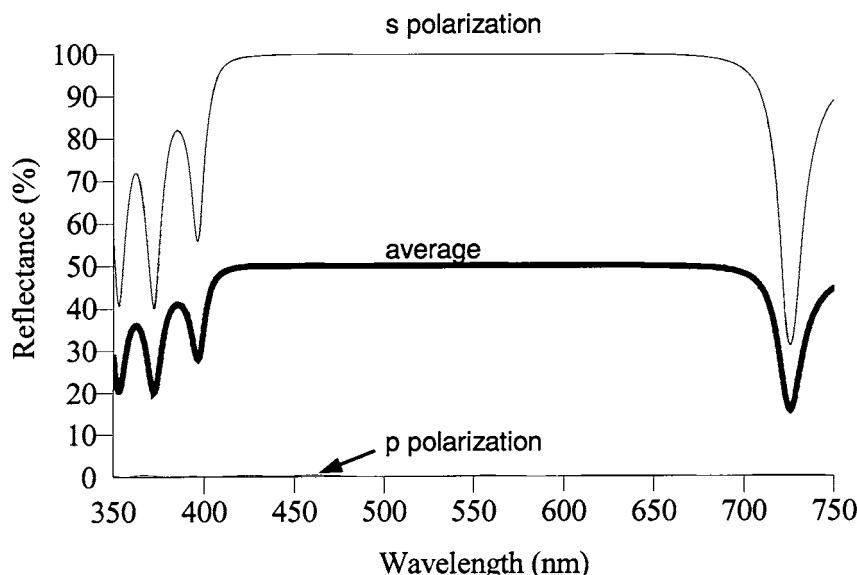
$$P_T = \frac{T_p - T_s}{T_p + T_s}$$

and in reflection

$$P_R = \frac{R_s - R_p}{R_s + R_p}$$

The *extinction ratio* indicates how well the polarizing beamsplitter discriminates between two planes of polarization. In transmission it is given as a ratio of T_p and T_s , and in reflection as a ratio of R_s and R_p . When the degree of polarization is very high, the reflected linearly polarized s-component and the transmitted linearly polarized p-component should each account for 50% of the incoming light intensity. Thus, an ideal polarizing beamsplitter acts as the 50/50 intensity beamsplitter, where each of the two emerging light beams are 100% linearly polarized (Fig. 20.10).

Figure 20.10
The Polarizing Polychromatic Cube Beamsplitter. The Computed Reflectance Represents a 15-Layer Design Consisting of Two Materials of High and Low Refractive Index. The Angle of Incidence is 52°



Optical Coating Process

Optical coatings are manufactured in high-vacuum coating chambers. Conventional processes require elevated substrate temperatures (usually around 300°C), whereas more advanced techniques, like ion-assisted deposition (IAD) are utilized at room temperatures. IAD processes not only produce coatings with better physical characteristics compared to conventional ones, but also can be applied to substrates made out of plastics. Figure 20.11 shows an operator in front of the optical coating



Figure 20.11

An Operator in Front of the Optical Coating Machine (*Courtesy of LaCroix Optical Co., Batesville, Arkansas*)

machine. Its main pumping system consists of two cryopumps. Control modules for electron-beam evaporation, IAD deposition, optical monitoring, heater control, pumping control, and automatic process control are in the foreground. Figure 20.12 shows the configuration of the hardware mounted on the base plate of a high-vacuum coating machine. Two electron-beam sources located at each side of the base are surrounded by circular shields and covered with shutters. The ion source is located in the middle. The optical monitor windows are in the front of the ion source. Figure 20.13 shows the upper part of the vacuum chamber,

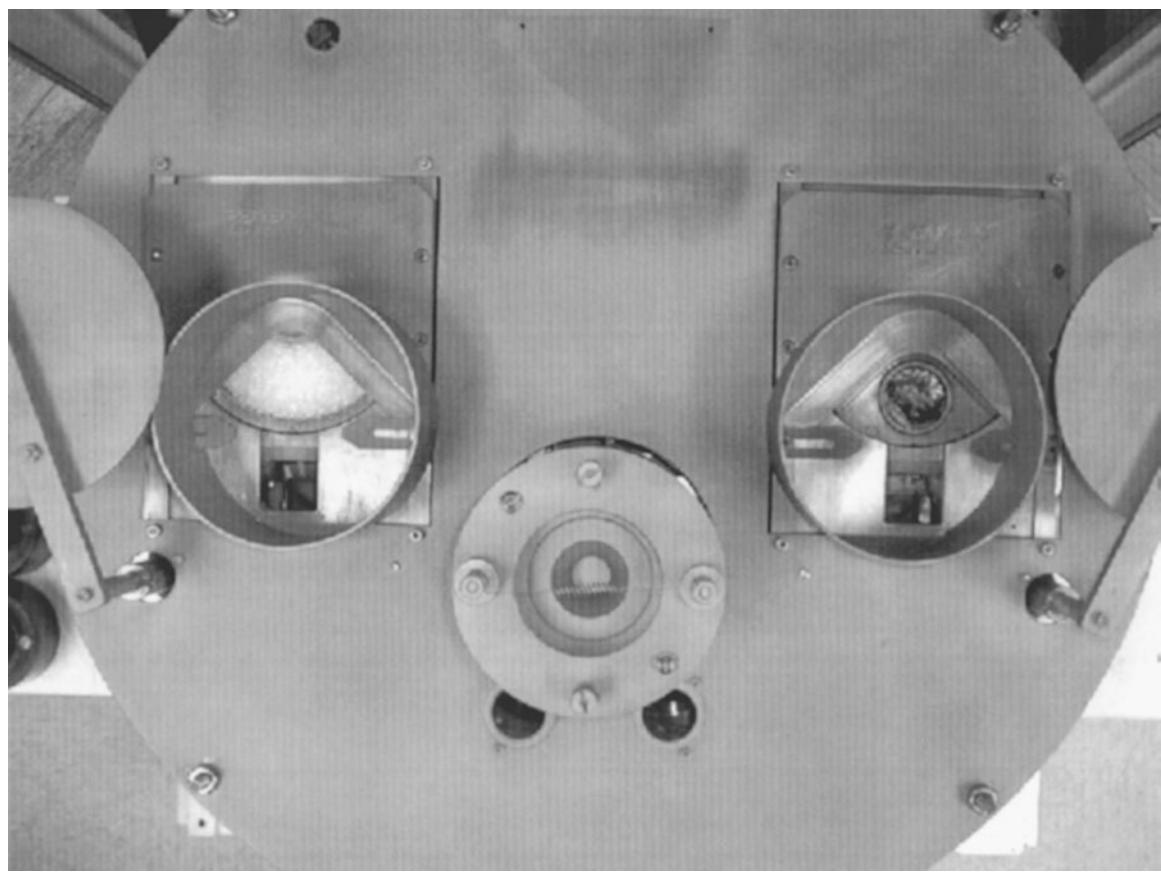


Figure 20.12

The Configuration of the Hardware Mounted on the Base Plate of a High-Vacuum Coating Machine (Courtesy of LaCroix Optical Co., Batesville, Arkansas)

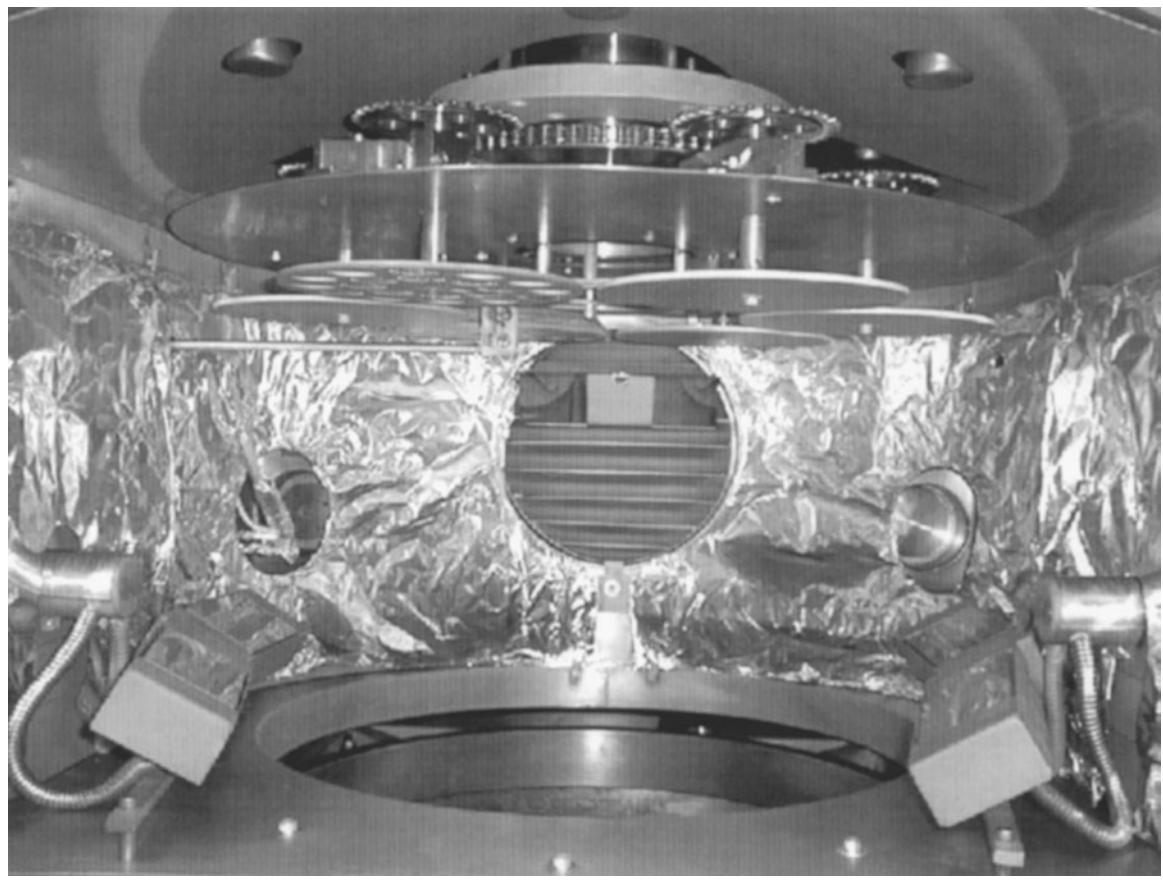


Figure 20.13

The Upper Part of the Vacuum Chamber Is Occupied by the Planetary System with Six Round Fixtures (Courtesy of LaCroix Optical Co., Batesville, Arkansas)

which is occupied by the planetary system with six round fixtures. Fixtures are loaded with optics to be coated. A use of a planetary system is a preferred way of maintaining a uniform distribution of the evaporated material across the area of the fixture. Fixtures turn around their common axis and revolve around their own axes. The optical and quartz monitors are in the middle of the planetary drive mechanism, the latter being obstructed by the drive hub. The large opening in the background leads to an additional high-vacuum pump. The substrate heating system consists of four quartz lamps, two at each side of the chamber.

The traditional methods of thin-film deposition have been thermal evaporation either by means of resistance-heated evaporation sources or by electron-beam evaporation. Film properties are determined mostly by the energies of the depositing atoms, which are only around 0.1 eV in conventional evaporation. IAD deposition results in direct deposition of ionized vapor and in adding activation energy to the growing film, typically in the order of 50 eV. Using the ion source, conventional electron-beam evaporation is improved by directing the flux from the ion gun to the surface of the substrate and growing film.

The optical properties of films, such as refractive index, absorption, and laser-damage threshold, depend largely on the microstructure of the coating. The film material, residual gas pressure, and substrate temperature can all affect the microstructure of the thin films. If the depositing vapor atoms have a low mobility on the substrate surface, the film will contain microvoids, which will be filled subsequently with water when the film is exposed to a humid atmosphere.

We define the *packing density* as the ratio of the volume of solid part of film to the total volume of film (which includes microvoids and pores). For optical thin films, it is usually in the range 0.75 to 1.0, very often 0.85 to 0.95, and rarely as great as 1.0. A packing density that is less than unity reduces the refractive index of evaporated material below the value of its bulk form.

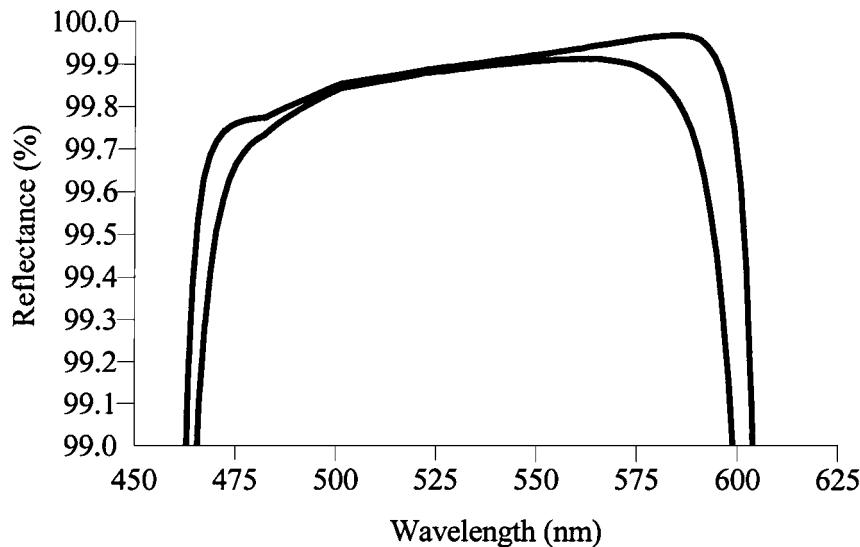
During the deposition, the thickness of each layer is monitored either optically or by using a quartz crystal. Both techniques have advantages and disadvantages that are not discussed here. What they have in common is that they are done in a vacuum while the material is evaporated. Consequently, they represent the refractive index of evaporated material in a vacuum, not the one that the material will acquire after being exposed to humid air. Moisture adsorption in the film results in displacement of air from microvoids and pores, causing an increase in the refractive index of the film. Since the physical thickness of the film remains constant, this refractive index increase is accompanied by a corresponding increase in optical thickness, which in turn results in the spectral shift of the coating characteristic toward a longer wavelength. To minimize this spectral shift caused by the size and overall population of microvoids throughout the growing film, high-energy ions are employed to convey their momentum to the atoms of evaporating material, thereby largely increasing their mobility during the condensation at the substrate surface.

Coating Performance Versus Number of Layers

We have mentioned earlier that the optical coating materials fall into two groups: *dielectrics* and *metals*. All of the preceding various optical coatings, except some metal mirrors, utilize dielectric materials in their design. Among dielectrics, the most often used are oxides and fluorides. One technological problem associated with the deposition of high-index oxide materials is their tendency to dissociate into oxygen and some lower forms of the original oxide. To avoid absorption in the depleted coating and to keep the coefficient A in Eq. (20.1) as close as possible to zero, it is necessary to *reoxidize* material before it condenses on the substrate, thereby preserving the *stoichiometry* of the bulk material.

One could think that the greater the number of layers, the better the coating performance. However, *given the manufacturing technology, there is a limit to a maximum number of layers that will produce the coating with the best characteristics*. For an optical designer just using an optical design program, it becomes a relatively straightforward conclusion that adding more surfaces and glasses to a certain, already well-corrected lens, for example, a double gauss photographic lens, will cause the image to deteriorate. Thin-film programs do not take into account physical characteristics of the coating microstructure and the atomic and molecular forces that exist between layers of different materials and within each layer. Consequently, thin-film programs cannot predict the physical behavior of the final coating design as much as the optical design programs can predict and characterize the image quality of an optical system. To illustrate this, let us take an example of a high-reflection dielectric mirror. It consists of a sequence of layer pairs of high- and low-refractive index materials (for example, TiO_2 and SiO_2), where each layer is QWOT. Assuming absorbing media, 12 of these layer pairs (implying a coating consisting of 24 layers) would boost reflectance to 99.9% at 530 nm. Adding another eight layers would not result in any considerable improvement. This is shown in Fig. 20.14. This 32-layer coating would have the same reflection of 99.9% at 530 nm, higher absorption, and greater overall thickness. Although with the slightly broader characteristic, it would probably be inferior to the 24-layer design because of a greater possibility of crazing (breaking off the coating because of high-tensile stress) and higher absorption that offsets the gain in reflectance.

Figure 20.14
The Reflectances of Two Dielectric Mirrors at 0° Angle of Incidence. Design Wavelength Is 525 nm, and the Coefficient of Absorption of High-Index Material Is 0.00027. The Upper Curve Represents a 32-Layer Design BK7| $(HL)^{16}$ |Air, and the Lower One Represents a 24-Layer BK7| $(HL)^{12}$ |Air. The Refractive Indices of Two Materials Are 2.336 and 1.461



Specifying Coating Requirements

Accurate specification of coating requirements assumes an understanding of the coating function, the function of the optical component to which it is applied, and the coating usefulness in a particular application.

For example, to increase the transmittance of an optical glass surface in the visible domain to 99.0% or more would require a broadband antireflection coating (BBAR) from 400 to 700 nm, for which Eq. (20.1) can be written in the following form

$$T = 1 - R - A \geq 0.99$$

or

$$R + A \leq 0.01 \quad \text{from 400 to 700 nm} \quad (20.2)$$

The last inequality expresses the requirement that the sum of the reflectance and absorptance should not exceed 1.0% for any wavelength in the interval 400 to 700 nm. Very often Eq. (20.2) is written as

$$R \leq 1.0\% \quad \text{from 400 to 700 nm}$$

assuming that the absorptance is close to zero ($A \approx 0$). If the glass is BK7 and the angle of incidence (AOI) of the light striking the glass surface is between 0 and 15° , then the fairly complete and accurately formulated requirement would be in the form

BBAR on BK7 glass

$$R(400 \text{ to } 700 \text{ nm}) \leq 1.0\% @ \text{AOI} = 0 \text{ to } 15^\circ$$

$$A \approx 0\%$$

To avoid some possible misinterpretations of the coefficient A , its maximum value can always be explicitly stated on the coating blueprint.

Relationship Between Production Cost, Tolerances, and Quality

The production cost per run of a particular coating is primarily determined by the size of the coating chamber, the manufacturing technology, and the complexity of the coating. Since the area of the coating chamber that can be used to coat parts is more or less directly proportional to the square of its radius, it follows that the bigger the chamber, the lower the price per coated lens. As an example, if the diameter of one chamber is twice the diameter of the other, then approximately four times more lenses can be coated in the first chamber than in the second one.

For some extremely stringent requirements, often found in the production of narrowband filters, it is not always possible to utilize the whole coating area within one chamber but rather one particular segment of it. This is because of the nonuniformity of the coating distribution across the chamber. Therefore, depending on the type of the coating, the capacity of the coating machine can be governed by the tolerances on the spectral characteristics of the coating.

For well-designed coating machines, the distribution of the spectral characteristic of evaporated material stays within $\pm 1\%$ of the nominal value. For example, the coating represented by Fig. 20.3 would have the range of reflectance minima from 495 to 505 nm. The inconsistency between different runs could further increase this range, say from 490 to 510 nm.

Besides the spectral conformity of the coated lens to the prescribed value, its quality is further governed by the least amount of coating voids, good adhesion and hardness, environmental stability, and the high packing density.

Different deposition techniques have been invented over the past 20 years in order to increase the packing density of evaporated material to the value close to unity. The most important ones are ion-assisted deposition (IAD), ion-beam deposition (IBD), and ion plating. We could finally say, the closer the packing density to unity, the more expensive the coating.

Bibliography

- Holland, L. (1956) *Vacuum Deposition of Thin Films*. London: Chapman & Hall.
- Jacobson, M. (1986) *Deposition and Characterization of Optical Thin Films*. New York: Macmillan.
- Macleod, H. A. (1986) *Thin-Film Optical Filters*. New York: Macmillan.
- Pulker, H. K. (1984) *Coatings on Glass*. Amsterdam: Elsevier.
- Rancourt, J.s D. (1987) *Optical Thin Films Users' Handbook*. New York: Macmillan.
- Thelen, A.d (1989) *Design of Optical Interference Coatings*. New York: McGraw-Hill.

This page intentionally left blank

Hardware Design Issues

There are many optical system design issues which relate directly to the ultimate hardware implementation, yet are different from the subjects we have covered thus far. It is important that the designer be reasonably fluent in these areas. They include the use of off-the-shelf optics, baffling and stray light control, and optomechanics.

Off-the-Shelf Optics

Off-the-shelf optics is, in effect, catalog optics. One of the significant advantages of off-the-shelf optics is that if what you need is in stock, you can have nearly immediate delivery. Unfortunately, the converse is also true: if what you need is *not* in stock, you may be faced with a long delivery time, perhaps in the order of 12 to 16 weeks.

The forms of off-the-shelf optics follow.

Precision Lens Assemblies

This first class of off-the-shelf optics includes relatively precision lenses such as camera lenses, relay lenses, enlarging lenses, and other multielement lens assemblies of reasonable quality. These lenses are most often mounted in nice-looking anodized housings, and may have adjustable f/numbers and focusing capability.

The optical and mechanical quality of these lenses may or may not be good. Just because the lenses are mounted in a beautiful black anodized housing with red, blue, green, and yellow engraving and just because the lenses are coated with a nice deep blue antireflection coating, there is no assurance whatsoever that the optical performance is any good. In addition, the focal length and *f*/number may or may not be per the specification. Moreover, the image quality may or may not be good. This is not to say that the specifications are not as advertised, nor is the performance necessarily poor; we only bring this up as a caution so that you are not misled by the external appearance of the lens assembly.

As with most off-the-shelf optics, these types of lenses are available almost immediately. If your lens is out of stock, delivery could take as long as 3 to 4 months, or longer, if indeed the same lens is ever again available. When you are dealing in a commercial commodity-like product line arena such as with 35-mm camera lenses, there is a rapid changeover in products, making future availability of a given lens a real questionable issue.

The cost of off-the-shelf optics in the form of completed lens assemblies can range from under \$100 to over \$500, or more, depending on manufacturing costs, volume, and, of course, quality. A good example of such a lens is a name brand 50-mm focal length *f*/2.0 35-mm camera lens, which we used recently for a laboratory test. It cost less than \$200 and performed extremely well for the intended purpose. The lens had six elements, and while its housing was partially plastic, it seemed robust enough for most applications.

Single Elements and Achromatic Doublets

The major catalog companies have several hundred different single-element lenses and achromatic doublets available. They typically range from approximately 1.5- to 2000-mm focal length, in diameters from approximately 1.5 to 150 mm. Their optical quality and level of tolerances are generally reasonable for many nondemanding applications and they are generally available uncoated as well as antireflection coated. Do not, however, expect to find extremely high-precision optics in this commodity area.

The cost of catalog single elements and doublets of small diameters up to approximately 50 mm are in the order of \$60 to \$150 each. The cost of *custom* single elements and achromatic doublets can be approximately \$350 per element in low quantities, and delivery can be 6 to 8 weeks. Delivery in 1 week is available from several vendors, naturally at a premium

price. As soon as the quantities increase to between 50 and 100, the price of custom lenses drops to prices close to catalog levels.

Other Forms of Off-the-Shelf Optics

There are many other forms of off-the-shelf optics available, including prisms, windows, mirrors, beamsplitters, polarization components, filters, and more. Also, there is the relatively new class of microoptics available off the shelf such as laser diode collimators and focusing optics.

How to Effectively Work with Off-the-Shelf Optics

If you are careful in use of off-the-shelf optics, you can be highly successful. On the other hand, if you are too casual and don't pay attention to details, your project could easily end up in trouble. Some guidelines gleaned over the years follow.

Complete lens assemblies are the most difficult to deal with. Manufacturers, such as the major camera companies, simply will not share with anyone the lens design prescription. Your ability, therefore, to input the design into one of the lens design and analysis software programs and interface it with other off-the-shelf, or even custom optics, becomes difficult, if not impossible.

If not given by the manufacturer, you could certainly have some of the basic parameters measured such as the focal length, *f*/number, and entrance and exit pupil locations. While not trivial to characterize, these parameters can indeed be measured. However, what you cannot do easily is measure the residual lens aberrations in order to factor them into a more complex system model to be used with other lens groups. Thus, incorporating off-the-shelf lenses with custom lenses can lead to serious problems.

There is one very important matter that must be considered, and that is that a lens designed for one set of specifications or parameters may or may not perform well under different conditions. For example, if we procure a 35-mm focal length *f*/2.8 double gauss camera lens from a well-known manufacturer, and we then proceed to use it at a near-unit magnification to image postage stamps or integrated circuit chips onto a CCD sensor, we will likely be very disappointed in its performance. The

lens was most likely designed for an infinite object distance or a distant object, and at a unity magnification, it will most likely perform very poorly. In addition to spherical aberration, the lens will suffer seriously from astigmatism and other off-axis aberrations.

The same holds for other specifications. Again, using this camera lens as an example, if we use it over a wider field of view or a larger spectral band than it was designed for, we will likely have poor performance. In addition, there may be distortion, which is not an image quality issue but rather a mapping error. If your application requires a precision machine vision lens with low distortion, then this must be measured for the proposed off-the-shelf lens.

Working with Off-the-Shelf Singlets and Doublets

This task is far more straightforward than working with complex lens assemblies due to two factors:

1. Many of the lens design software packages have included the design prescriptions of singlets and achromatic doublets from most of the major catalog suppliers. For example, Zemax has resident lens prescriptions from Edmund Scientific, Melles Griot, Opto-Sigma, Rolyn, Newport Corporation, Coherent, Spectra Physics, and Linos Photonics. Fortunately, the suppliers of these lenses realize that they can serve the technical community far better by providing this information rather than being secretive.
2. The lenses are fundamentally simple lenses with little to be concerned about with respect to pupils for example. In addition, even if the prescription is not available, you could generate a candidate design and have a moderate level of confidence that the real lens will be close to the catalog lens.

For example, let's assume that you find in some new catalog an achromatic doublet which has a focal length of 78 mm and a diameter of 10 mm, but the catalog is not resident in your design package. If you are confident that the lens was designed for an infinite object distance, you could in a matter of a few minutes emulate it with reasonable confidence of the design being at least sufficiently close to the actual design to be useful in your computer modeling. You might, for example, select

BK7 glass for the crown element and SF2 glass for the flint and optimize it. The results should be reasonably close to the real lens.

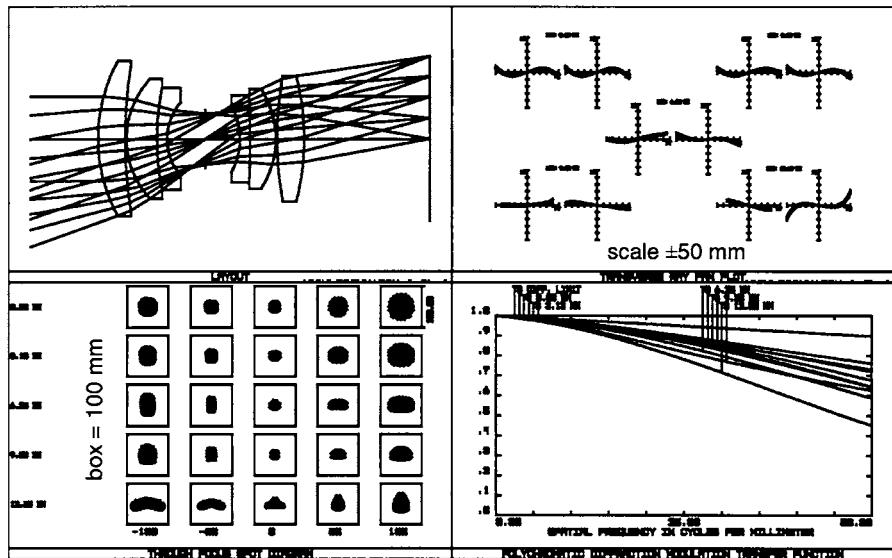
In developing a lens design for which you intend to explore the potential of using off-the-shelf singlets and/or doublets, a good procedure to follow is to first perform the design yourself so as to meet the system's first-order and performance specifications. You may want to begin with a first-order design using so-called paraxial lenses, and later convert it to a real design. Once you feel comfortable with your design, then you need to evaluate the focal length of the singlets and/or doublets which you intend to match to off-the-shelf components. If you intend to use a planoconvex singlet, then in the design you should also use a planoconvex element; the same holds true for planoconcave lenses and equiconvex or equiconcave lenses. Now you need to look in one or more catalogs for lenses that match closely the parameters of your lenses (in particular, the focal length and diameter) and replace your lenses with the catalog lenses. Most of the software packages allow you to simply insert any off-the-shelf lens into an otherwise custom design. Make sure you pay attention to the lens orientation, or which way the crown and flint elements are oriented.

At this point, you may find that your performance and other specifications are adequately met, in which case you can freeze the design and procure the lenses. On the other hand, you may find that for one reason or another the design requires further optimization, in which case you need to comply. This may require customization of one or more lens groups for example. Often your final design might include a mix of off-the-shelf components as well as custom components. You will likely find that as you incorporate more off-the-shelf components into a given design, its performance will degrade from optimum. However, the important question to be answered is whether the performance is *good enough*.

Example of Lens Used at Conjugates Different from What It Was Designed

To illustrate some of the preceding issues, Fig. 21.1 shows the layout and performance for a 35-mm focal length $f/2.8$ double Gauss lens designed for an infinite object distance. Figure 21.2 shows the performance of the same lens with an object distance of 0.5 m. Note that the plot scales are maintained and are identical in all of the figures in this analysis, and the

Figure 21.1
A 35-mm Focal Length $f/2.8$ Lens at Infinity



modulation transfer function (MTF) data are plotted to 50 line pairs/mm. At first glance, the transverse ray aberrations look similar to the previous nominal design data, and indeed there is not a significant degradation. However, note that the MTF has suffered a significant drop, especially off axis. In Fig. 21.3, we show the same lens at a 100-mm object distance (this results in a demagnification of $3\times$). The performance is significantly

Figure 21.2
A 35-mm Focal Length $f/2.8$ Lens at 500-mm Object Distance

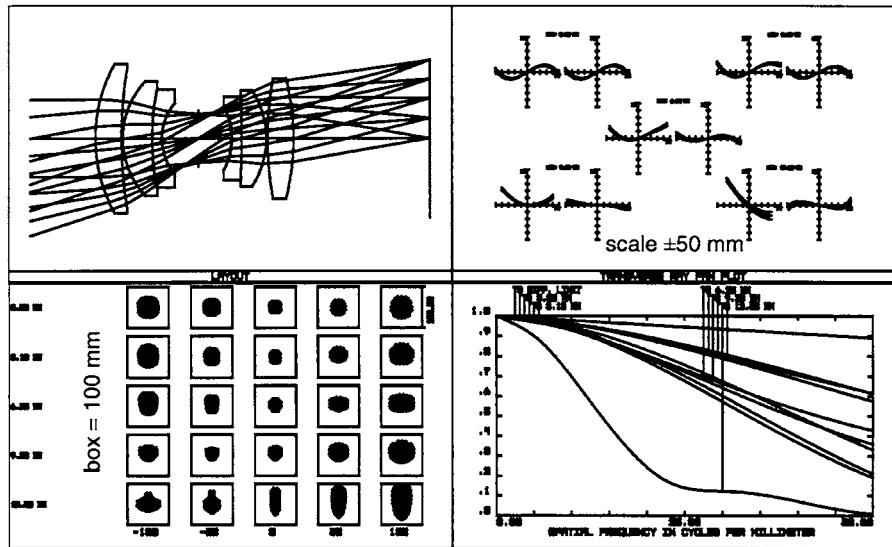
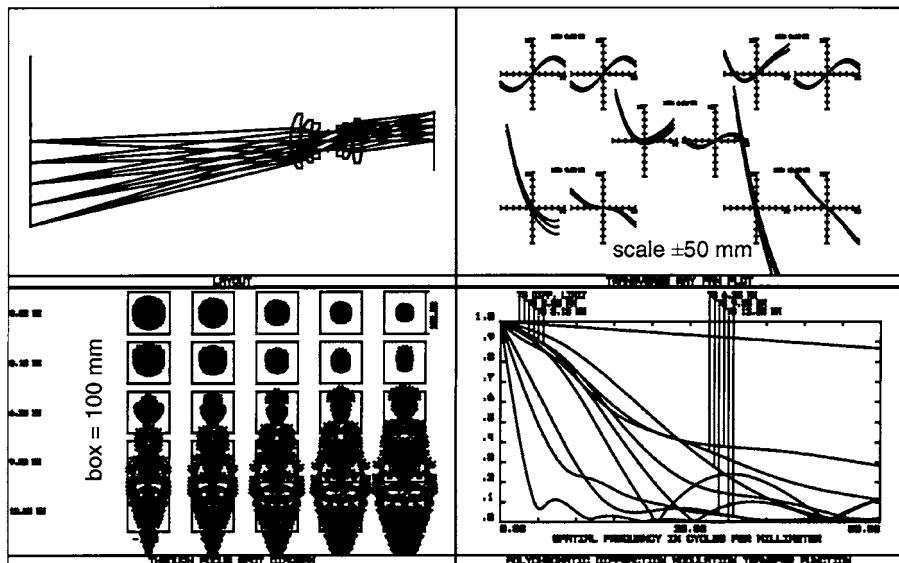


Figure 21.3
A 35-mm Focal Length $f/2.8$ Lens at
100-mm Object Distance



degraded from the nominal lens performance. Figure 21.4 shows the performance if the lens is used at a 1:1, or unity, magnification. In this case, the performance is extremely poor. At the edge of the field there is over an order of magnitude increase in spot diameter when the lens is used at 1:1 magnification!

Figure 21.4
A 35-mm Focal Length $f/2.8$ Lens at Unit Magnification (32.41-mm Object Distance)

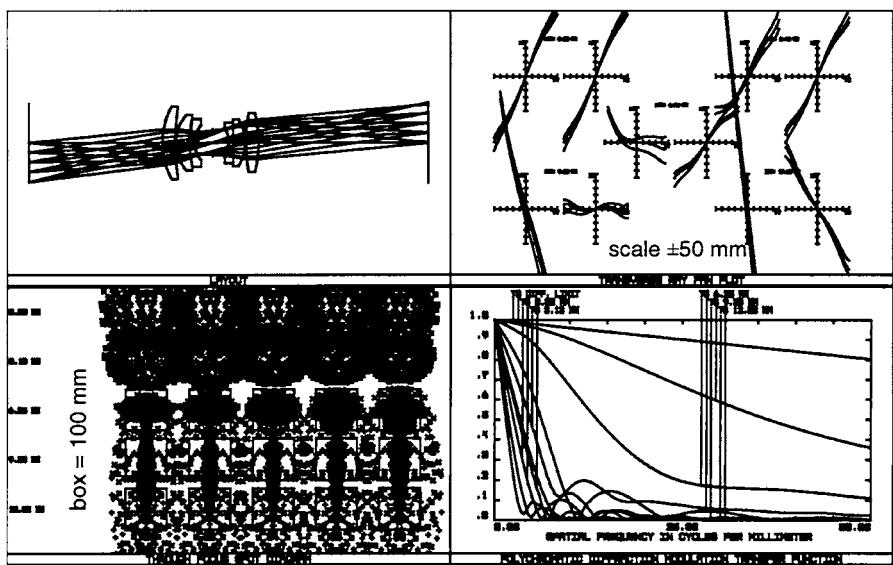


Figure 21.5
RMS Blur Diameter
for a 35-mm Focal
Length $f/2.8$ Lens As
a Function of Object
Distance

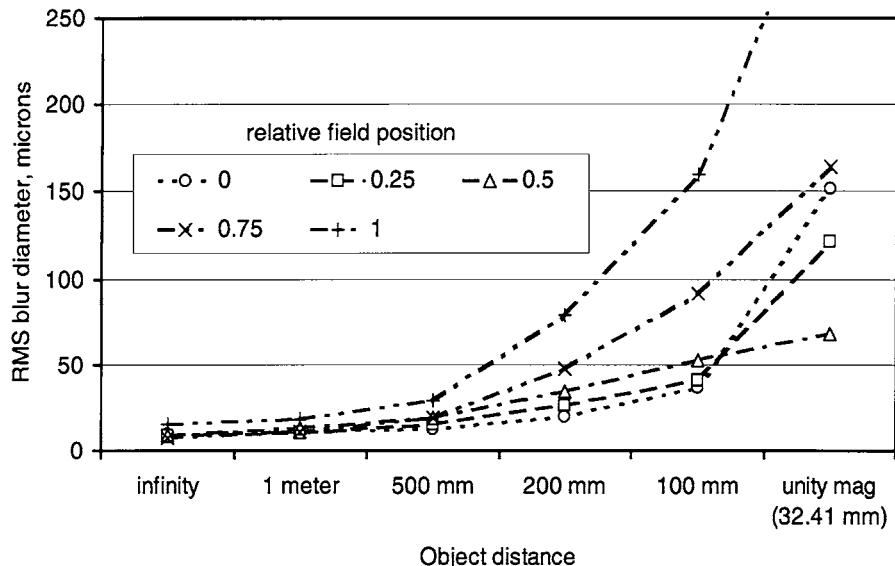


Figure 21.5 shows parametrically how the performance degrades as a function of object distance for the previous lens design example. Note that the nominal design gives an rms blur diameter of approximately 9 μm over most of the field of view. This doubles for the 0.5-m object distance, and for a 100-mm object distance the rms blur diameter increases to about 50 μm over the central region of the field of view. At the unit magnification position, due to the extreme aberrations introduced, the spot diameter ranges from 70 to about 340 μm at the edge of the field. If you were using a CCD chip with a 12- μm pixel pitch, an object distance of no more than 0.5 to 0.75 m would be viable in order to maintain a reasonable modulation at the Nyquist frequency.

Pupil Matching

In addition to the basic specifications, performance, and aberrations, the extremely important issue of entrance and exit pupils must be considered when working with off-the-shelf optics. Clearly, if you were simply using an off-the-shelf lens to image an object onto a CCD array, the location of the pupils is of little concern or interest. However, if you had a multiple-stage relay system, then the exit pupil of one stage must be coincident or

nearly coincident with the entrance pupil of the next stage, and so on. As we learned earlier, field lenses are indispensable in this task, as one of their primary roles is to reimagine the exit pupil of one lens group into the entrance pupil of the next lens group. This issue can be of major concern when one or more of the off-the-shelf lenses are zoom lenses, since the pupils can translate or move over great distances as the lenses zoom, and having a mismatch in pupils is very likely, if not inevitable.

Development of a Lab Mockup Using Off-the-Shelf Optics

There are many situations where determining the level of performance of your system quickly is to your advantage. Situations where this approach is helpful are when validating important aspects of your specifications. For example, assume you are designing a new visual telescope. Parameters, which are important, include field of view, magnification, eye relief (clearance from the last element to the eye), and of course image quality. You could manufacture a prototype of your custom production design, which would likely take several months and cost many dollars. An alternate approach is to build up a unit using off-the-shelf optics. It should be straightforward to nearly meet the magnification and field-of-view specifications, and likely the eye relief too. The image quality may be degraded from your custom production design; however, the overall ability to assess the general nature of the system specifications and performance is often quite valuable. This is especially true for some of the specifications, such as field of view and magnification, which may have been based on a judgment or best-effort basis. You can take your mockup system outside and use it in a near-real functional environment. There is always an anticipated level of performance associated with every lens design, and your system performance can often be demonstrated using off-the-shelf optics.

Stray Light Control

The suppression of stray light is often ignored until it is too late, and then it becomes costly and time consuming to fix the problem. Good engineering in this area is imperative. The best way to learn the subject

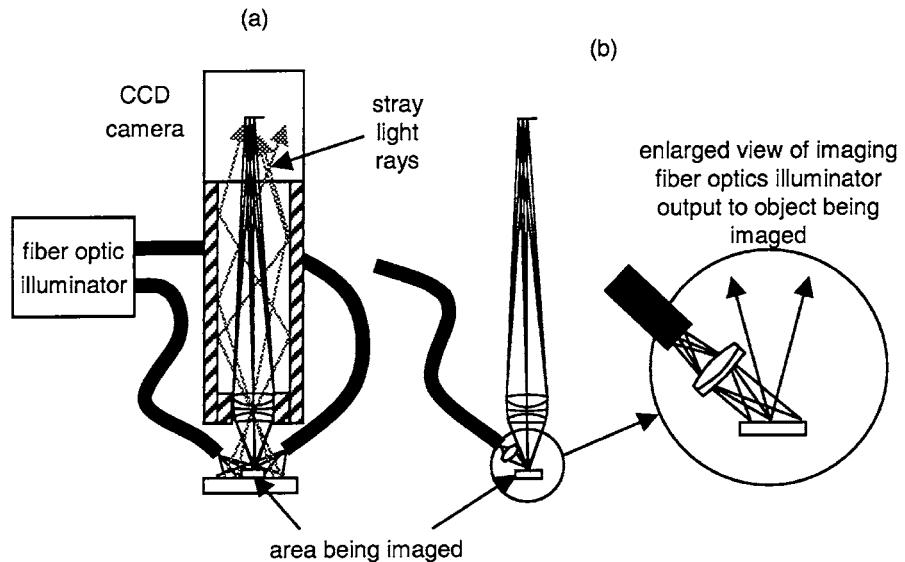
is through the following two examples: (1) a machine vision system and (2) a reflective Cassegrain telescope.

Machine Vision System

We will first relate a true story regarding a potentially serious stray light problem:

- We were called in to visit a colleague who said that he had just installed a new vision system and the contrast was badly degraded from prior systems. The system was very basic and consisted of a microscope objective and a CCD camera. The contrast was indeed poor on the video monitor.
- We first removed the camera from the tube assembly and looked in with our eye at a location similar to where the CCD chip was. It was immediately evident that there was a lot of stray light reflected from the interior of the tube assembly. While the tube interior was black anodized, at near-grazing angles of incidence black anodizing is quite reflective. Figure 21.6a shows the situation.

Figure 21.6
Example of Stray Light Control in Machine Vision System



- We then asked our host if he had any flat black paper, and he did. We rolled the paper into a tube shape and put it into the anodized tube assembly. Visually, with our eye again looking into the tube, we found that the situation was indeed improved but not perfect.
- After reinstalling the camera, our host said, “wow, that’s a lot better... but it isn’t as good as it used to be.” This was consistent with our observation.
- We then took a careful look at the overall system layout and we realized that the fiber-optic illuminators were illuminating an area far exceeding the object being imaged, as shown in Fig. 21.6a. We asked our host if he had a small positively powered singlet or doublet lens and he did. We then cut a small aperture in a piece of black paper and fastened it to the end of the fiber bundle. The lens was now used to reimagine the aperture onto the object being imaged, as shown in Fig. 21.6b. A little experimentation with the magnification resulted in a situation where we were just overfilling the object area of interest.
- We now turned the system on and our host said, “wow, that’s better than it has ever been!”

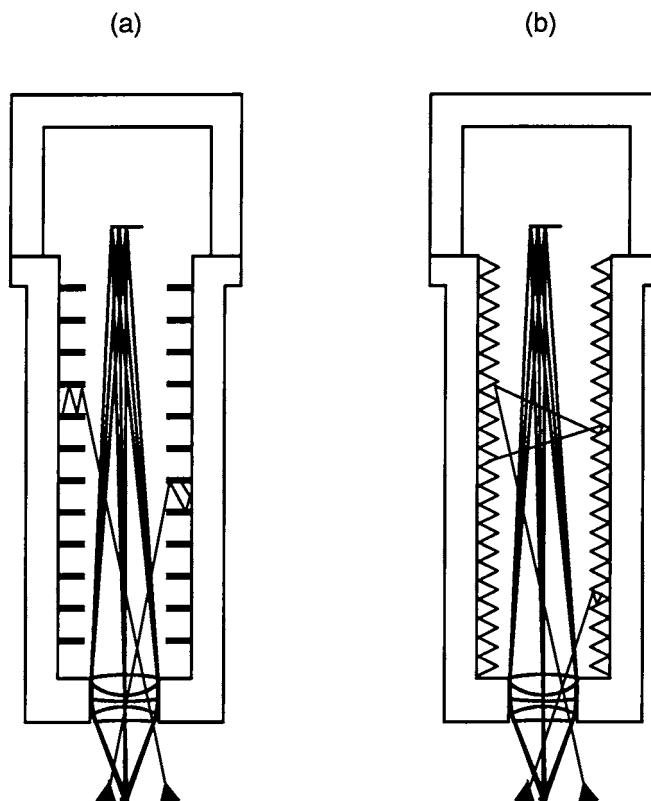
If you were to design the microscope tube assembly for optimum stray-light attenuation, it would be best to incorporate baffle features on the interior of the tube, as shown in Fig. 21.7. Figure 21.7a shows a series of baffle structures similar to washers. This is one of the most efficient baffle forms; however, it is somewhat costly to machine or otherwise implement. Figure 21.7b shows a coarse thread with which we can derive good results. Note that we show the multiple bounce path of several representative rays, and in the case shown none of the rays reaches the CCD sensor until after three bounces, which is a good guideline. Do keep in mind that there will inevitably be scattering and diffraction coming from the tops of the threads, no matter how perfectly they are machined, and you will be better off with a coarse thread rather than a fine thread with more thread tops. One final tip: it will help if you make the inner diameter of the tube and associated baffles as large as possible.

Cassegrain Telescope

One system that always requires efficient stray-light baffling is the Cassegrain telescope, which was discussed in Chap. 8. Without baffling, there is generally a direct stray-light path from the object space to the

Figure 21.7

Use of Threads and
Baffles for Stray-Light
Attenuation in
Machine Vision
System



sensor, and that could be a serious problem. We generally use two basic baffles, one a conical baffle extending aft from the edge of the secondary mirror along the limiting imaging ray bundle, and the other a tubular baffle extending forward from the hole in the primary mirror.

In order to show how to baffle a Cassegrain, we first generated a candidate design. We selected an $f/8$ system with a 100-mm entrance pupil diameter covering a full 1° field of view. The goal for our baffles is that a limiting ray that just passes by the two baffles described earlier shall not directly strike the image plane. In order to quickly and efficiently reach a solution, we added a central obscuration to the computer model and traced 500 rays into the entrance pupil at each field of view. Figure 21.8a shows the model. Areas in black are fully populated with rays, and the clear regions extending aft from the secondary mirror and forward from the primary mirror are available for baffles.

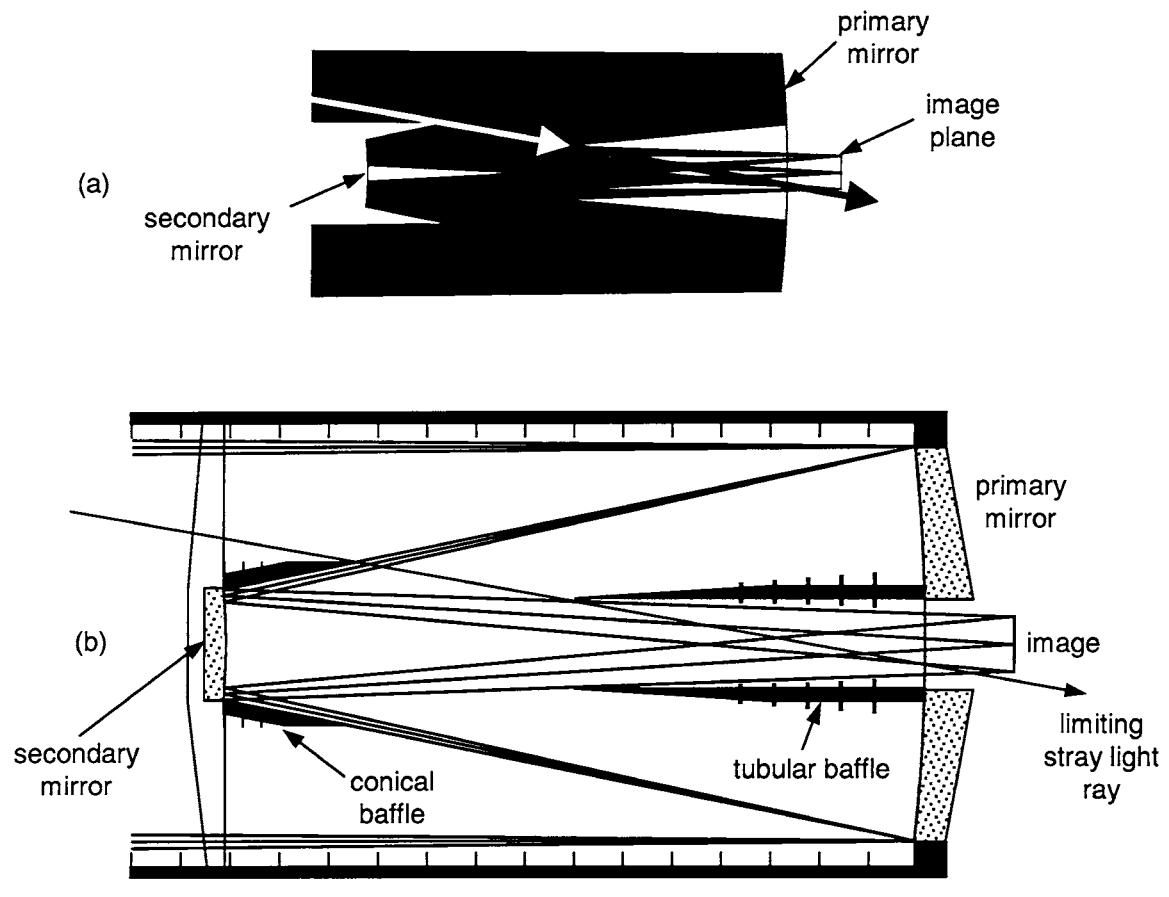


Figure 21.8
Stray-Light Baffling of Cassegrain Telescope

We show also the limiting ray which just clears the two baffle ends and reaches the image plane.

Figure 21.8b shows an implementation of this baffle. Note that we have added vane-type baffle segments as presented earlier in this chapter. We have also added an outer-tube assembly with interior baffling.

It is important to realize that good common sense and a little dedicated work will generally provide you with efficient stray-light baffling. If you need a specific attenuation factor, then you will need to use one

of the stray-light software packages. For example, in space applications, where a system may be observing a black sky to within a few degrees of the Sun, stray-light attenuation in the order of 10^{-15} or more is often required. We showed earlier in Fig. 8.7a configuration especially well suited for efficient stray-light attenuation. This system is a three-mirror configuration consisting of primary, secondary, and tertiary mirrors. Let us assume that we are in space looking within a few degrees of the Sun into a black sky. There will be a large amount of light scattered and diffracted from the edge of the primary mirror since it is receiving direct solar radiation. If we now locate a stop further aft in the system at an image of the primary and slightly reduced in size from the image of the primary mirror, we will effectively block this light from proceeding further through the system. This is known as a *Lyot stop* after the French astronomer Lyot. While Fig. 8.7 is not to scale, it does illustrate the principle involved.

Optomechanical Design

The design of the mechanics to support your imaging optics is extremely important. Design issues relating to the optomechanics are the following:

- The mechanics supports the lenses and/or mirrors in the system. In order to keep the image quality within the specification, every optical component must be held to its nominal position within the required tolerances, as derived from your tolerance analysis and system performance error budget.
- The mechanics, along with the optics, must perform over the required thermal environment. The designer must allow for thermal expansion and contraction of the optical components to prevent any catastrophic problems.
- Maintaining focus through temperature is very dependent on the optics as well as the mechanics, and athermalization may be required.
- The mechanics must fit within the desired packaging space and be within its weight goal.
- The mechanics must aid in attenuation of stray light. This is often accomplished by blackening the housing interior as well as threading and providing stray light baffles at strategic locations.

Figure 21.9 shows a typical housing for a projection lens. In use, a reflective display device is located to the left of the light-injection prism. The image generated on the display is then projected to a screen to the right of the lens system.

We have pointed out some of the important mechanical design features to incorporate into the design. These are

- In this design the aperture stop is between the two smaller elements, and we use a spacer as a physical aperture stop.
- The elements are held in place by a front-threaded retainer and a rear retainer.
- A thread is applied to the conical spacer between the two left-most powered lens elements. This threading is to attenuate any stray light which may be incident on the housing.

We show for reference a different lens housing in Fig. 21.10. Note in this lens the two left elements are bonded into the housing as evidenced

Figure 21.9
Typical Lens Housing

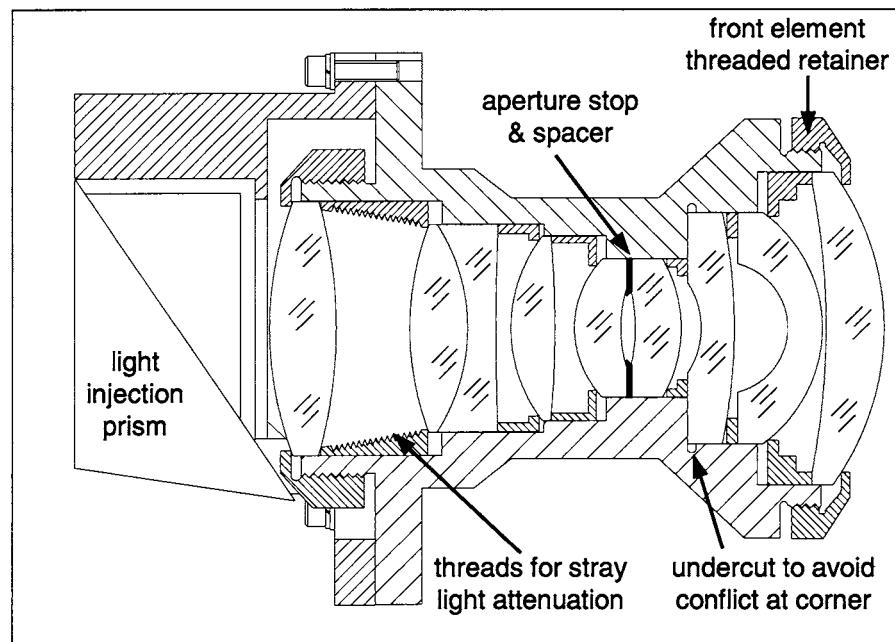
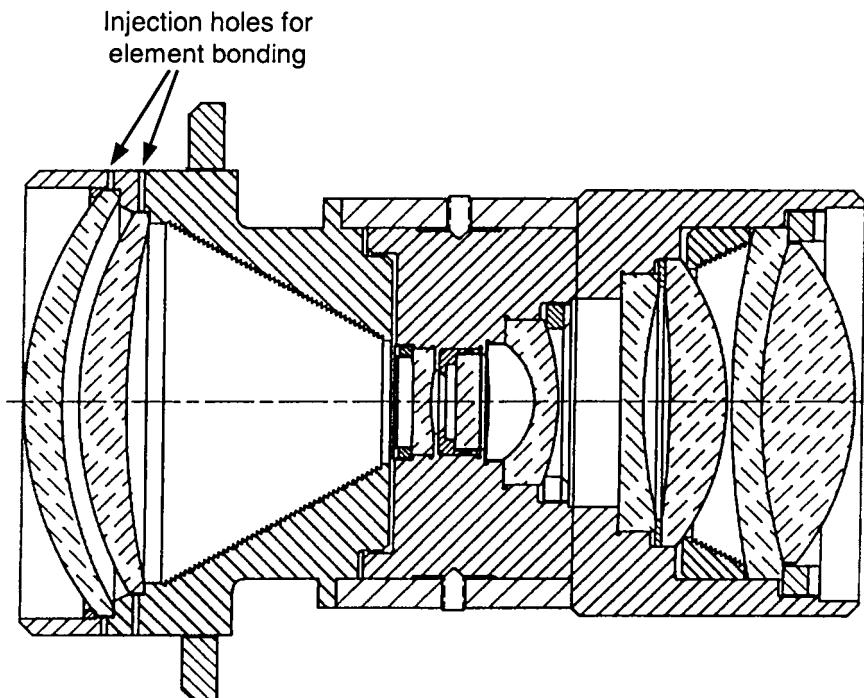


Figure 21.10
Typical Lens Housing



by the two material-bond injection holes. The bond material is typically a semicompliant epoxy or RTV. This is done in situations where shock and vibration may be a problem. The elements are centered using shims or by rotating the housing on a precision air bearing and assuring that the runout of the housing and the elements are per the tolerance callout.

Lens Design Optimization Case Studies

In this chapter we will guide you through several representative design studies and parametric analyses in order to demonstrate the design process. We will begin with the design, from basic principles, of an achromatic doublet. Included will be the detailed computer input and output using the Zemax software package, one of the industry's standards. Following a successful design effort on the doublet, we will show the design of a low f /number double Gauss lens similar to a high-quality 35-mm camera lens. And then we will work through a case study for a digital camera lens. Following this, we will show the design for a 7×50 binocular. And, finally, we will show a parametric analysis of single-element and achromatic doublets using various manufacturing technologies, including aspherics, diffractive surfaces, and others.

Error Function Construction

Prior to embarking on several design examples, we need to discuss how the measure of performance, or the *error function*, is computed in a lens design program. Since this error function must be computed a very

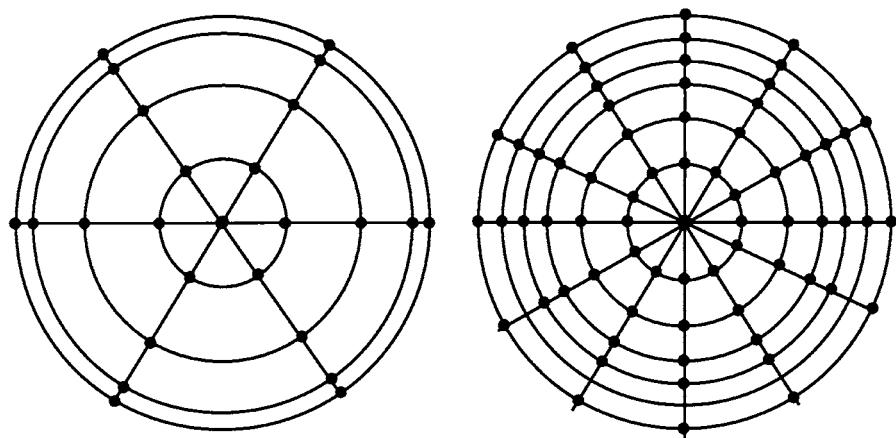
large number of times during the optimization process, it must be kept as simple as possible so it computes quickly. The construction of an error function was discussed in Chap. 9.

We could use the third-, fifth-, and seventh-order aberrations for our error function. These are very fast to compute; however, with today's complex systems, these aberrations rarely represent sufficiently well the real performance. We could alternatively combine these third-, fifth-, and seventh-order aberrations with specific ray aberrations or optical path differences at selected fields of view and entrance pupil coordinates. This approach can solve the problems of the higher-order aberration residuals; however, there is a lot of user interaction involved, which makes this a user-intensive methodology.

Perhaps the best and easiest to use error function is the rms blur diameter at the image formed by a grid of rays traced into the entrance pupil. The error function could also take the form of the rms wavefront error, or other similar criteria. Regardless of which method is used, it is specified at each wavelength and at each field position along with appropriate weightings.

The grid of rays in the entrance pupil is shown in Fig. 22.1, where we show on the left the default grid of three rings and six arms. Specifically, the rays traced are at the intersection of the rings and arms. On the right we show a denser grid formed by 6 rings and 12 arms. The default

Figure 22.1
Ray Grid of Three Rings and 6 Arms (Left) and 6 Rings and 12 Arms (Right)



3 rings, 6 arms, 18 rays/color

6 rings, 12 arms, 72 rays/color

grid represents 18 rays in the entrance pupil per color per field position, and the denser grid represents 72 rays in the entrance pupil per color per field position. Denser grids are used when higher-order aberrations are present so as to better sample the aberrations. This is often the case when aspheric surfaces are used, for example. Overall, you must consider constantly whether you are sampling the rays or OPDs sufficiently well in the entrance pupil, the fields of view, and the wavelengths. If not, more rays, more fields, and/or more wavelengths are required. The computer really doesn't care what grid density is used; it will minimize the ray or OPD aberrations specifically at the grid points you specify, and *only* at those points.

The merit function is a numerical representation of how closely an optical system meets a specified set of goals. The *operands* in the merit function represent not only the image quality but also focal length, magnification, size constraints, etc.

Achromatic Doublet Lens Design

The specifications for our doublet are shown in Table 22.1.

To begin, we will derive a simple achromatic doublet so as to have a decent starting point for the computer optimization. The *V* number, or Abbe number, of optical glass is

$$V\# = \text{Abbe}\# = \frac{n_d - 1}{n_F - n_C}$$

TABLE 22.1

Achromatic
Doublet Design
Example

	Parameter	Specification
	Entrance pupil diameter (mm)	50.8
	Focal length (mm)	254
	<i>f</i> /number	<i>f</i> /5
	Full field of view (degrees)	±2
	Spectral range (μm)	Visual (C, d, F) (0.6563, 0.5876, 0.4861)

where n_f is the refractive index at 0.486 μm or shorter wavelength and n_c is the refractive index at 0.6563 μm or longer wavelength. Further, we showed in Chap. 6 for lens elements a and b that

$$\phi = \phi_a + \phi_b \quad \phi_a = \frac{\phi V_a}{V_a - V} \quad \phi_b = -\frac{\phi V_b}{V_a - V_b}$$

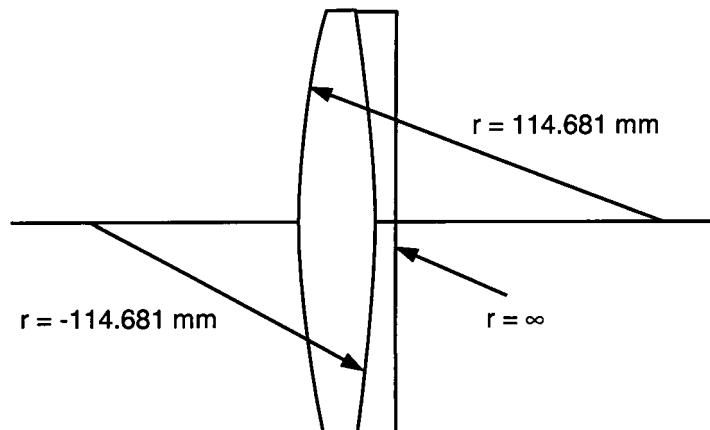
We will assume that the positive crown element is BK7 glass with a refractive index $n_d = 1.517$ and a dispersion $V_a = 64.5$, and for the negative flint element we assume SF2 glass with a refractive index $n_d = 1.620$ and a dispersion $V_b = 36.3$. Based on these glass assumptions, we find that $\phi_a = 0.009$ (focal length = 111 mm), and $\phi_b = -0.0051$ (focal length = -197.1 mm). Further, let us assume that the positive crown element is equiconvex and the negative element is planoconcave, and the elements are cemented. For a thin lens we have shown that

$$\phi = \frac{1}{f} = (n - 1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right)$$

For the crown which is to be equiconvex, $r_2 = -r_1$ and we derive the radius to be $r = 114.681$ mm. For the negatively powered planoconcave flint element we find that $r_2 = \infty$ and $r_1 = -122.806$. We will cement the two elements, as shown in Fig. 22.2.

We now show in Fig. 22.3 the lens prescription data on the top and the optimization data on the bottom. The following points relate to the lens data input:

Figure 22.2
Achromatic Doublet
as Input to Computer



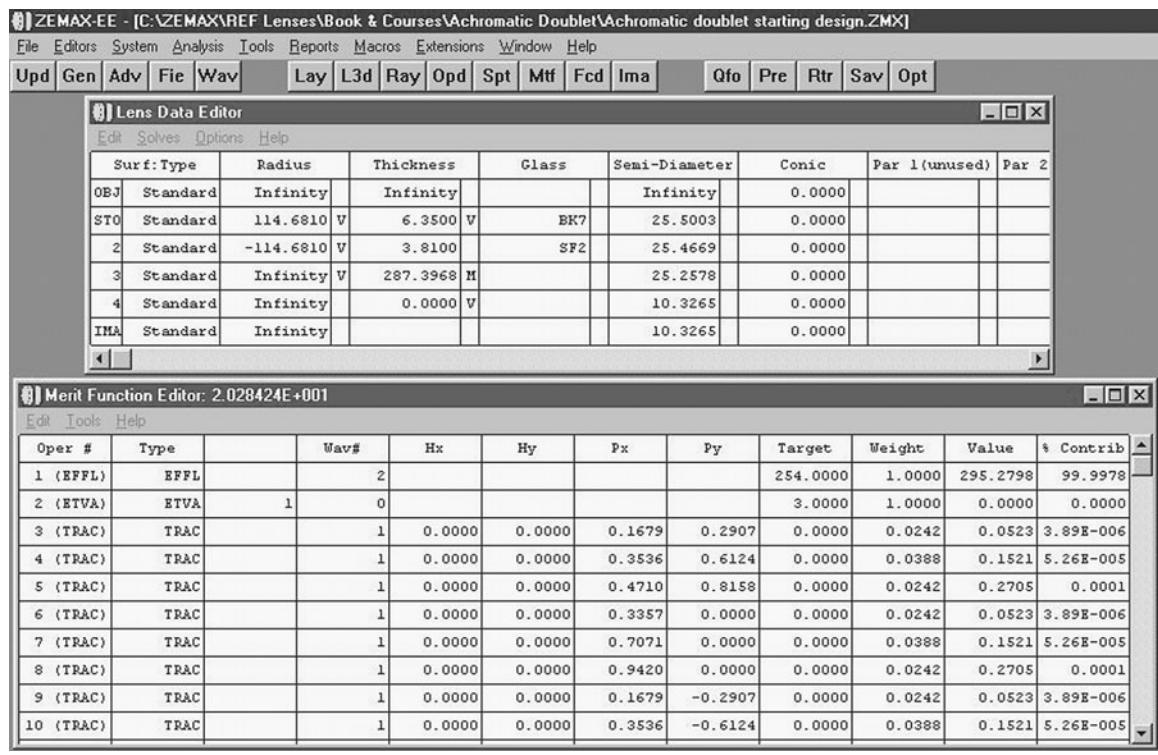


Figure 22.3

Achromatic Doublet as Input to Computer

- The data are input in spreadsheet format.
- The radii of 114.681 and -114.681 are input in the radius column.
- These two radii, as well as the rear flat surface, are assigned the letter "V" which indicates that the radii are variable in the optimization.
- The thicknesses are assigned reasonable values. Surface 3 (surface numbers are on the far left) is designated an "M" to the right of the thickness. This means that the thickness is to be the distance to where the paraxial ray height equals zero (287.3968 mm).
- The aperture stop is on surface 1. Surface 1 is also variable in thickness in order to allow a reasonable edge thickness for the element. Note that if you do control the edge thickness of an element, you *must* vary its thickness.

- The glasses are listed as appropriate on surfaces 1 and 2.
- Separately in what are known as the “General,” “Field,” and “Wavelength” editors in Fig. 22.4, we input the entrance pupil diameter, the fields of view, and the wavelengths with their associated weights.
- We add a surface (number 4) which we vary independently in thickness. This is the refocusing from the paraxial focus to the best focus position. Using this technique can sometimes result in a better-controlled optimization process.

In the lower part of Fig. 22.3 we show the optimization window, and we note here the following:

- The first line is labeled “EFL” which means the *effective focal length*. Our target is 254 mm and its current value is 295.2798. We assign this a weight of unity in the optimization.
- The second line is “ETVA” which is the *edge thickness value* on surface 1, the positive first element. Our goal here is 3 mm. From line 3 onward, we have as labeled “TRAC” the transverse shift from the ideal image locations in the image plane. There is a very *important* side note here: Recall that we have varied the thickness of the first element (surface 1). If we had not varied the element thickness but did require the constraint of the edge thickness of the element to be 3 mm we would be overconstraining the lens. The program would reach a solution; however, we would, in effect, be constraining the power of the first element by this edge thickness constraint, and the net result would be a poorer level of optical performance.

Figure 22.5 shows the initial lens design and its performance. The transverse ray aberrations are plotted on a 200- μm scale and the spot diagrams are on a 500- μm scale.

We now execute the optimization, and only a few seconds later a local minimum in the error function is reached. Figure 22.6 shows the design data as well as the error function construction, only here for the optimized doublet. Figure 22.7 shows the layout and the performance of the design. The transverse ray aberrations are now plotted on a scale of 50 μm which is 25% of the initial scale factor, and the spot diagrams are on a scale of 100 μm , which is 20% of the initial scale. The lens is clearly better now than the starting design.

If we required a further performance improvement, we would likely introduce a small airspace between the two elements and possibly

Figure 22.4
 Means for Entering
 Basic System
 Parameters, Including
 Entrance Pupil
 Diameter, Field
 Angles, and
 Wavelengths

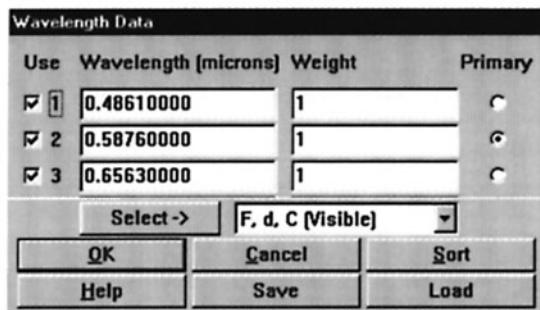
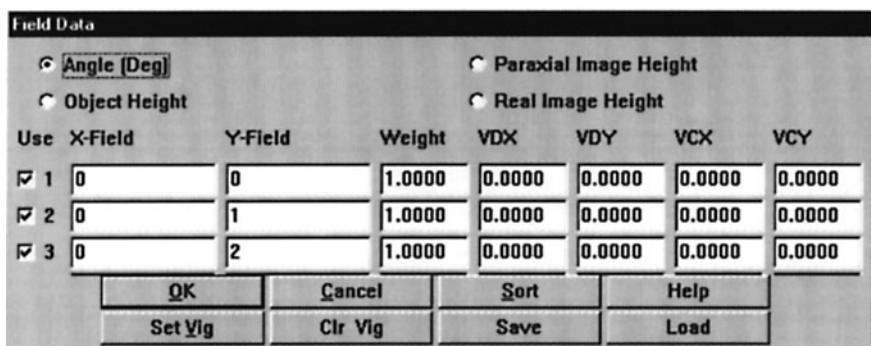
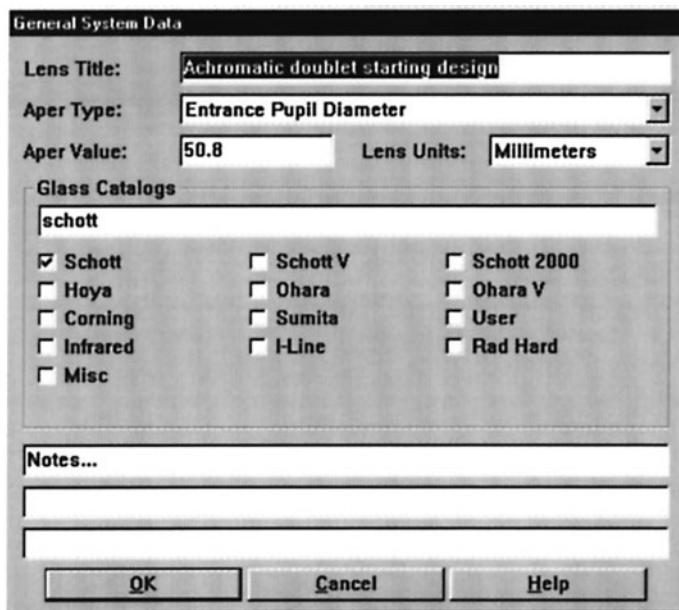
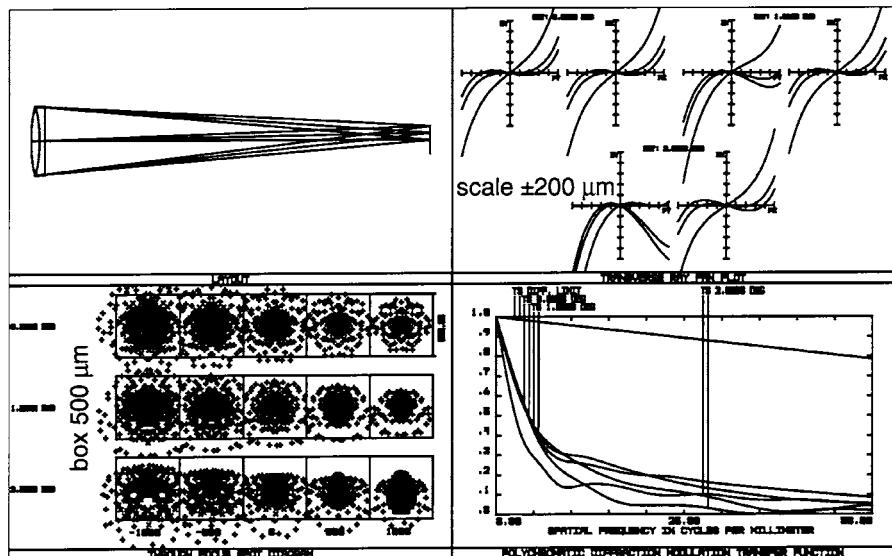


Figure 22.5
Performance of
Achromatic Doublet
as Input to Computer



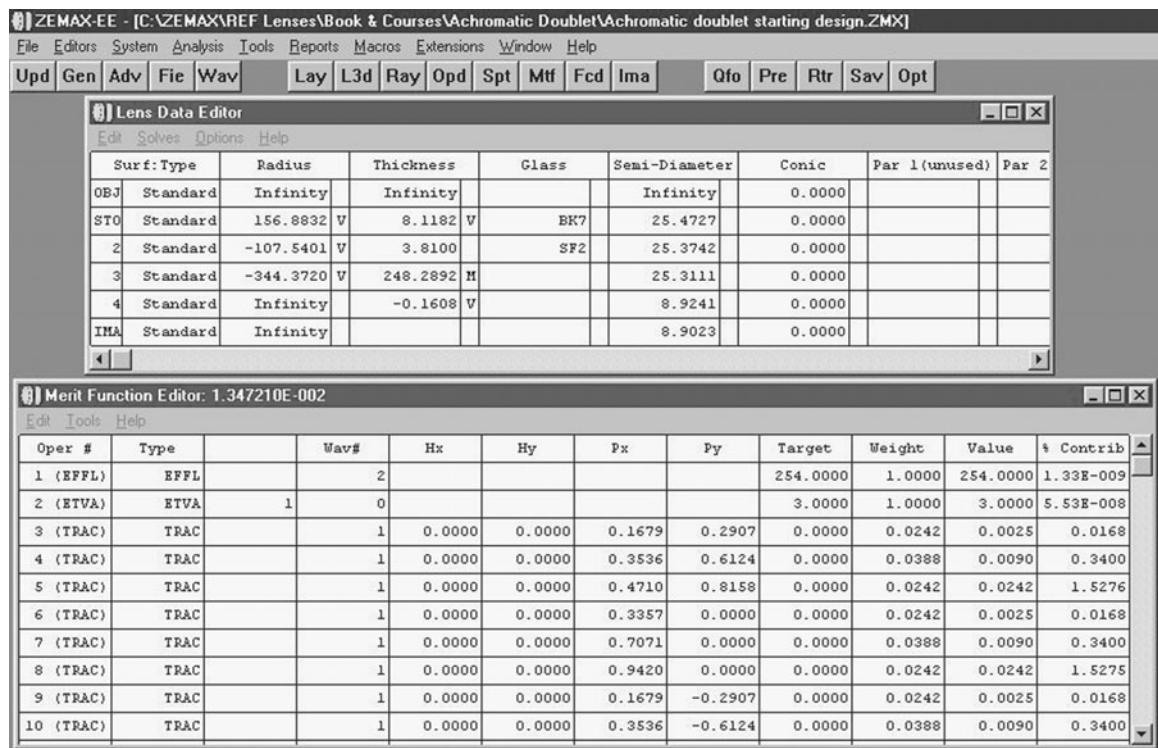
change the glass types. The airspace would allow for a balancing of third- and fifth-order spherical aberration which can often make a significant difference, and the glass change would allow for improved chromatic aberration correction. A further improvement would be realized by adding a third element. If this new element were near the focal plane, it would be able to minimize both the field curvature as well as the astigmatism. We leave these exercises to the reader.

Double Gauss Lens Design

This example is the design and tolerancing of a 50-mm focal length double Gauss lens for a 35-mm camera application. The basic lens specifications are shown in Table 22.2.

Relative to the optical performance, we will derive our own specification based fully on the functional performance requirements of our 35-mm camera lens. We will then show how this derivation comes extremely close to what is often used in the industry.

Let us assume that the goal for lens resolution is that the image blur from a point object is barely discernable by the eye when viewed on a 200- × 254-mm enlargement at a distance of 254 mm. This is a pretty

**Figure 22.6**

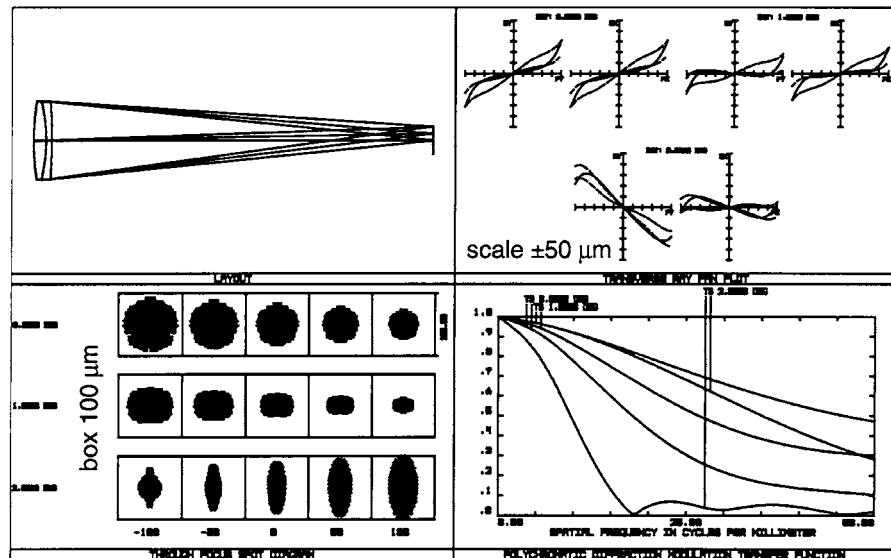
Design Data and Error Function for Achromatic Doublet Final Design

reasonable specification, and the beauty is that we did not need to know any complex aberration theory to come up with it. It is a functional specification based solely on the application of the lens, which is, of course, to give the user a good quality photograph.

Figure 22.8 shows the situation. A person with good visual acuity can resolve about 2 min of arc per line pair, which equates to 1 min of arc per line, which is approximately 0.0003 rad. At a distance of 254 mm, this equates to an image blur diameter of 0.076 mm. (For reference, 1 min of arc is a spot 1 mm in diameter at a distance of 3 m.) A 35-mm negative which measures 24 × 36 mm is 7.06 times smaller than the enlargement, which means that we are looking for an image blur diameter of 0.0107 mm on the negative.

Prior to getting into our design example, consider the common rule of thumb that a 35-mm camera lens should have an MTF of >0.3 at

Figure 22.7
Performance of
Achromatic Doublet
Final Design



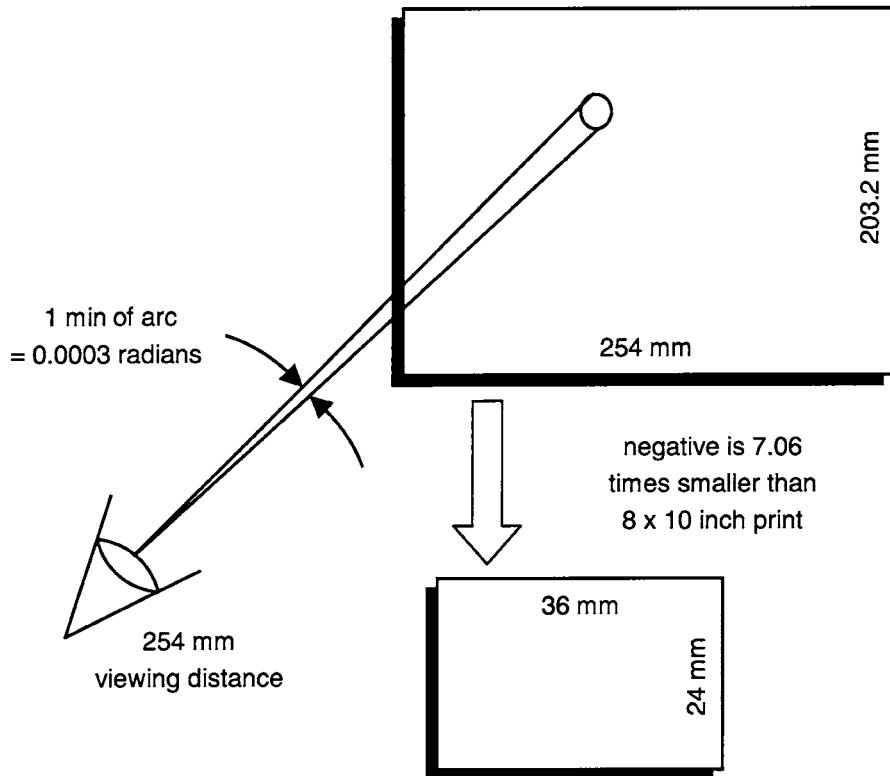
50 line pairs/mm and an MTF of >0.5 at 30 line pairs/mm. We can rederive this guideline as follows: Our image blur diameter goal is 0.0107 mm at the lens focal plane. This is approximately equivalent to a line $\frac{1}{100}$ mm wide. A line pair, which is a dark line and an adjacent bright line, is therefore $\frac{1}{50}$ mm wide, which equates to 50 line pairs/mm. This very interestingly matches our rule of thumb perfectly. We could then conclude that a reasonable contrast level for such a lens would be an MTF of about 0.3 at 50 line pairs/mm, exactly what our rule of thumb calls for!

TABLE 22.2

Double Gauss
Design Example

Parameter	Specification
Entrance pupil diameter (mm)	25.4
Focal length (mm)	50.8
f/number	f/2.0
Full field of view (degrees)	±16
Spectral range	Visual (C, d, F)
Distortion (%)	<2.5
Vignetting (%)	<50 at edge of field
Back focus (mm)	>25.4

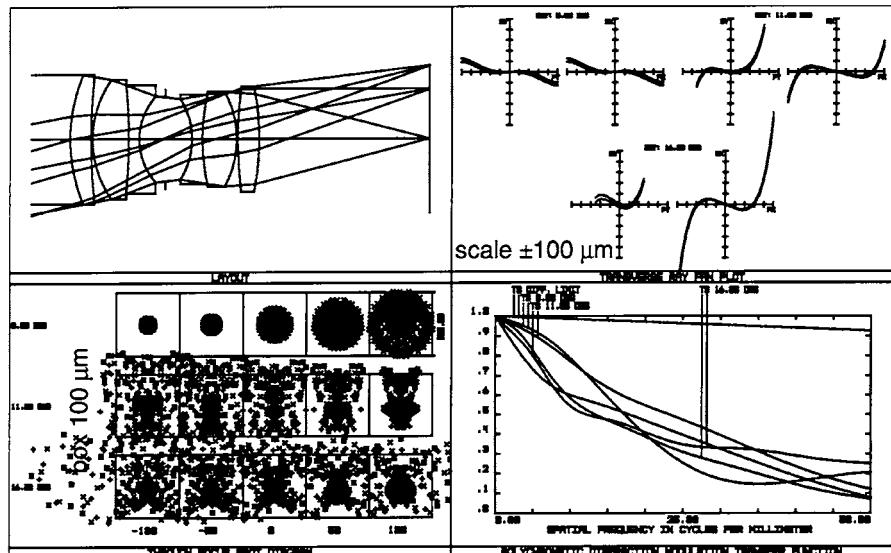
Figure 22.8
Performance Derivation for Double Gauss Lens



Let us now select as a starting point a double Gauss lens design from a 1938 patent. After setting up the prescription on the computer, we have the results shown in Fig. 22.9. These data include a lens drawing or layout, a plot of the transverse ray aberrations, a through-focus geometrical spot diagram, and a plot of the MTF out to 50 line pairs/mm. The error function is a combination of different constraints, the ray aberrations, and other performance criteria. Our constraints include the focal length, the edge thickness of the positive elements, the minimum back focus distance, and the distortion. The pure lens quality portion of the error function is 0.018, and it is this metric that we will be following as we optimize the lens.

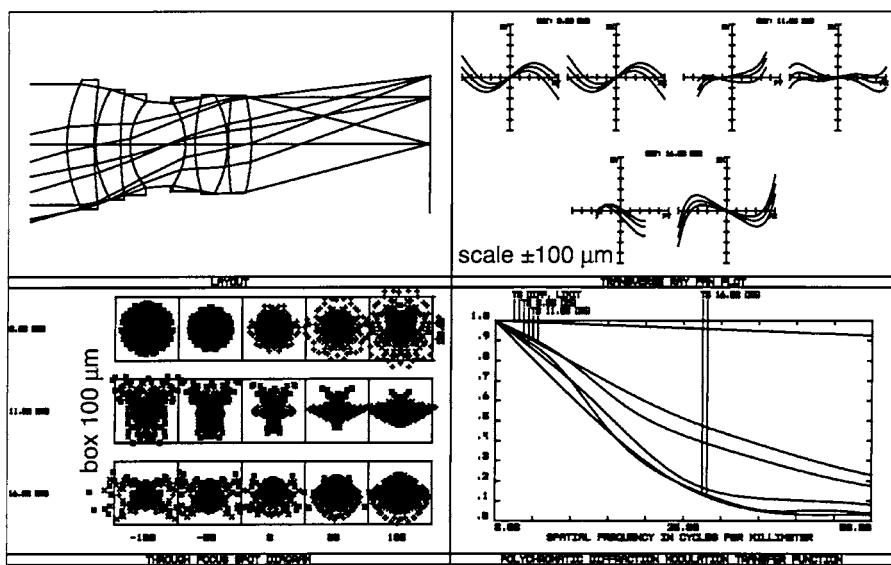
Step 1. We now take the initial patent lens prescription and establish our variables. Variables are all of the lens radii, the airspaces surrounding the aperture stop, the thicknesses of the positive elements, and the back focus distance from the rear lens

Figure 22.9
Performance of
Double Gauss
Starting Design
from Patent



vertex to the image plane. Figure 22.10 shows the result of this initial optimization. We have a spherical aberration residual, primary axial color, and field curvature, among other residual aberrations. The error function has reduced from 0.018 for the starting design to 0.0106.

Figure 22.10
Performance of Initial
Optimization



Step 2. We now will vary the inner doublet glasses using a routine called “Hammer” optimization in Zemax. The Hammer optimization uses a random search algorithm in the solution space surrounding the starting design. The program will, by definition, end up with real glasses, so the user does not need to be concerned with fictitious glasses after a long optimization. We allowed the Hammer optimization to run approximately 30 min, enough time to realize a moderate improvement in the lens performance. Figure 22.11 shows the results. The spherical aberration is somewhat reduced and the chromatic aberration is nearly eliminated. The error function has reduced from 0.0106 to 0.0060, another reasonable improvement.

Step 3. We now allow all of the glasses to vary, including those of the outer elements. Figure 22.12 shows the results, and the error function reduces from 0.0060 to 0.0059, only a slight improvement. Note here that the MTF is about 0.15 at the corner of the field of view in the tangential target direction at 50 line pairs/mm. Our goal is 0.3 minimum at the edge of the field of view, so we have a way to go.

Step 4. Up to this point, we have been optimizing and analyzing our design at the center of the field of view, 0.7 of the

Figure 22.11
Vary Inner Doublet
Glasses, 30-min
Hammer

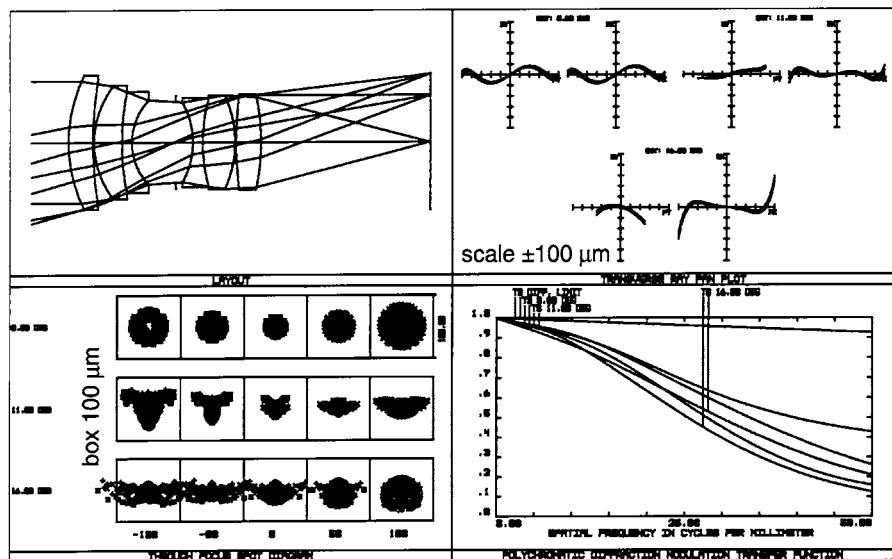
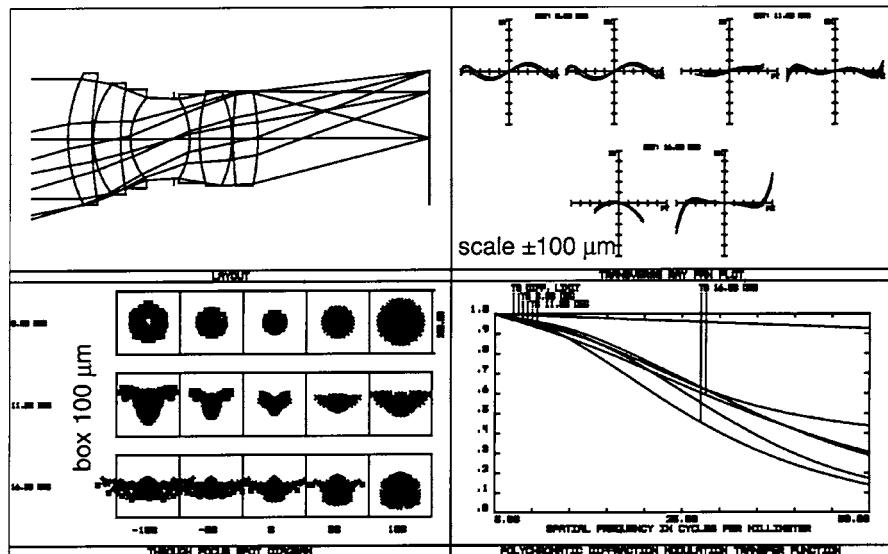


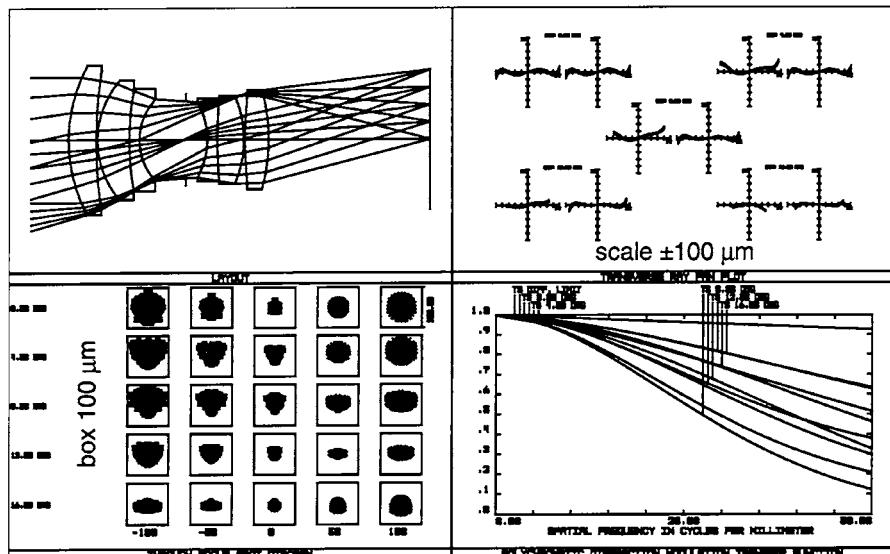
Figure 22.12
Vary All Glasses



semidiagonal of the field of view, and at the corner which is the maximum field of view. This is often an adequate sampling of the fields of view, especially when there are not significant changes in performance with field. But we have a reasonably low f /number lens with a reasonably wide field of view, so we elected to increase the number of fields of view to 5 in equal increments. Figure 22.13 shows the performance. Something extremely noteworthy has happened, and that is the inner fields of view really do not perform well! Note that at 25 and 50% of the field of view our rms blur diameter is 25% larger than at the corner of the field of view! This is evident in the ray trace plots as well as the spot diagrams. We actually did two other things in this model: We defined the aperture diameters on each surface, and we then adjusted the vignetting factors so that realistic vignetting would result. Unfortunately, if the user does not go through this exercise, the vignetting may be fictitious and not representative of what will happen in hardware. The net result of this was that the higher-order flare in the sagittal ray fan at the corner of the field of view was eliminated which improved the off-axis performance. This was partially

Figure 22.13

Five Fields, Set Apertures Then Vignetting, No Reoptimization



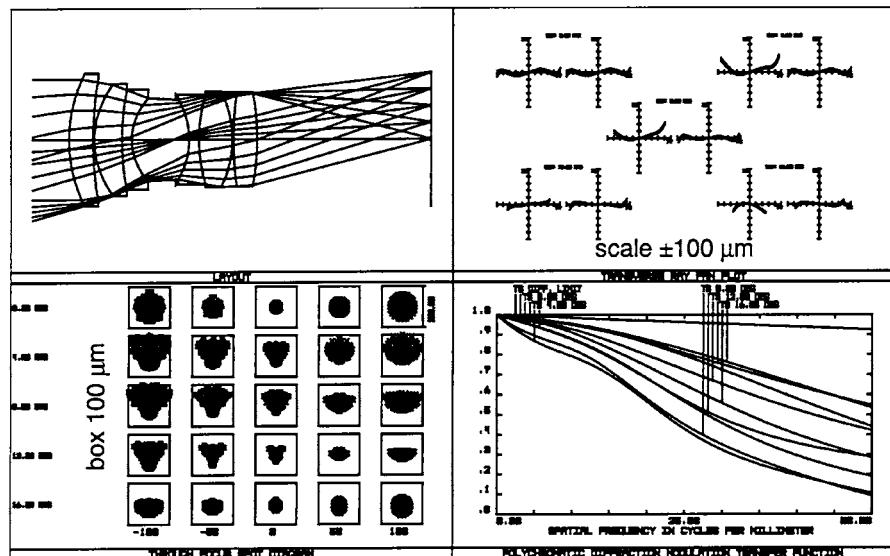
responsible for the error function coming down from 0.0059 to 0.0047, even though there was no reoptimization of the lens in this step.

Step 5. In the next step we applied the basic optimization algorithm and Fig. 22.14 shows the result. The error function comes down from 0.0047 to 0.0042. The inner fields of view still show more degradation in performance than the center or corner of the field, and we will need to do something about this problem. We could increase the field weights at these positions; however, as you will see, this was not necessary.

Step 6. In our next iteration we continued to use the increased field sampling of five fields of view, and further we increased the ray sampling in the pupil to 6 rings and 12 arms in order to assure an adequate sampling. While this was not mandatory, we do have some higher-order aberrations, and it makes good sense to increase the sampling at about this stage in the design. Figure 22.15 shows the result. While the error function only reduced from 0.0042 to 0.0038, the performance at the inner fields of view are clearly improved. We should note here that any time you change the ray sampling via the number of rings and arms, the field and/or spectral weights, or other similar parameters, you need to recompute your error

Figure 22.14

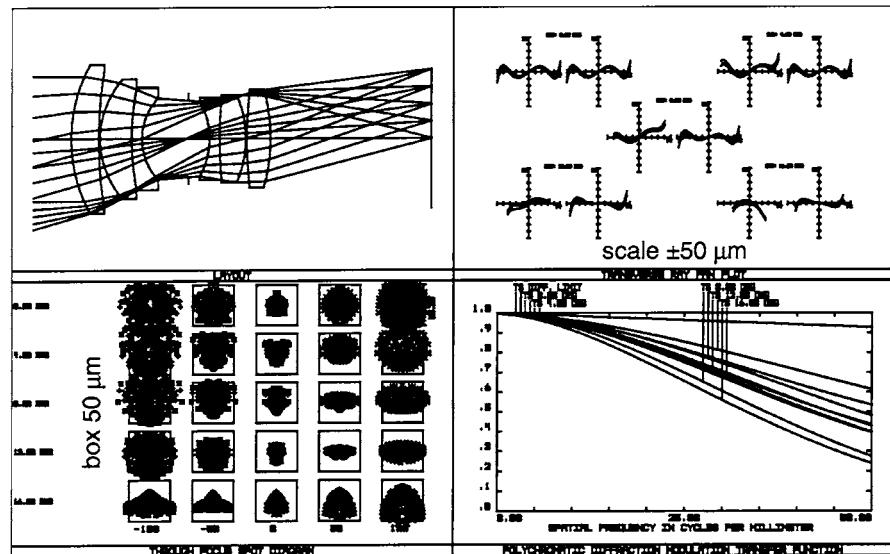
Five Fields, Set Apertures and Vignetting,
Basic Optimization



function. Note also that the new error function will likely be different from what it was before due to the different sampling, changes in field and/or wavelength weights, or other factors, which have changed. And this is true even if the lens itself is unchanged.

Figure 22.15

Five Fields, Tighter
Ray Grid (6 Rings, 12
Arms), Basic
Optimization



Step 7. In the next iteration we allowed all glasses to vary one final time and executed a 30-min Hammer optimization. The error function reduced slightly from 0.0038 to 0.0034, and the MTF actually seemed to degrade somewhat from the prior design, so we are not showing the results.

Step 8. In the final iteration we allowed the Hammer optimization to run for a full 12 h, and a much improved design resulted, as shown in Fig. 22.16. The error function has come down from 0.0034 to 0.0023, approximately a 30% reduction. The ray trace curves, spot diagrams, and MTF all show a notable improvement in both basic performance as well as uniformity of performance over the field of view. Note that the lowest MTF is now 0.6 at 50 line pairs/mm! The rms blur diameters range from 6 to 9 μm over the full field of view.

Figure 22.17 shows a Pagel diagram of the final design of our $f/2$ Double Gauss lens. This shows us which surfaces introduce significant primary aberrations such as spherical aberration, coma, astigmatism, and others. For example, surface 8 just after the stop introduces the most spherical aberration and as expected surfaces 6 and 8 introduce the most field curvature which balances the field curvature of the other surfaces

Figure 22.16
Final 12-h Hammer
Optimization

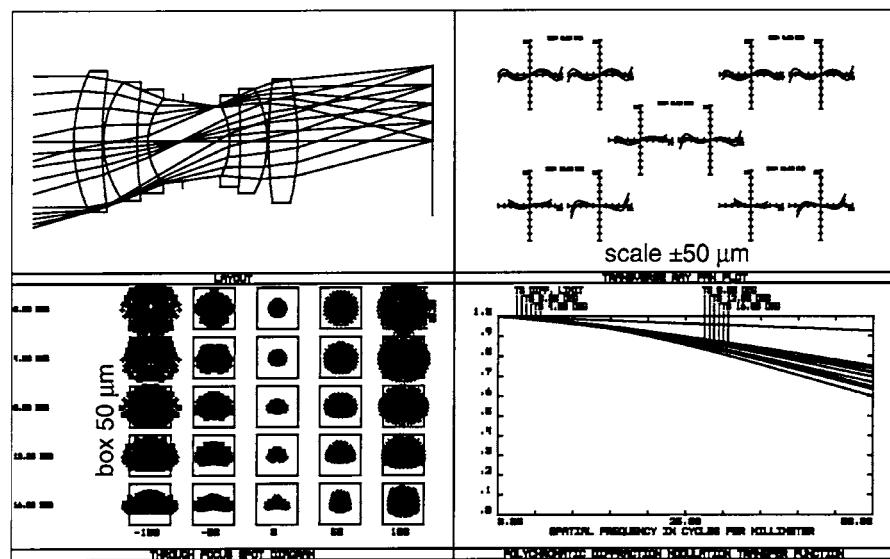
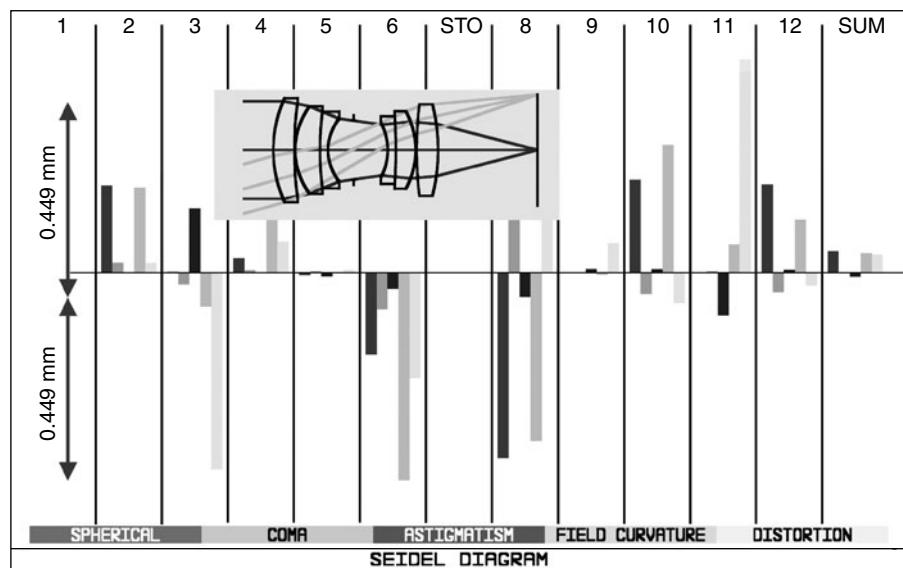


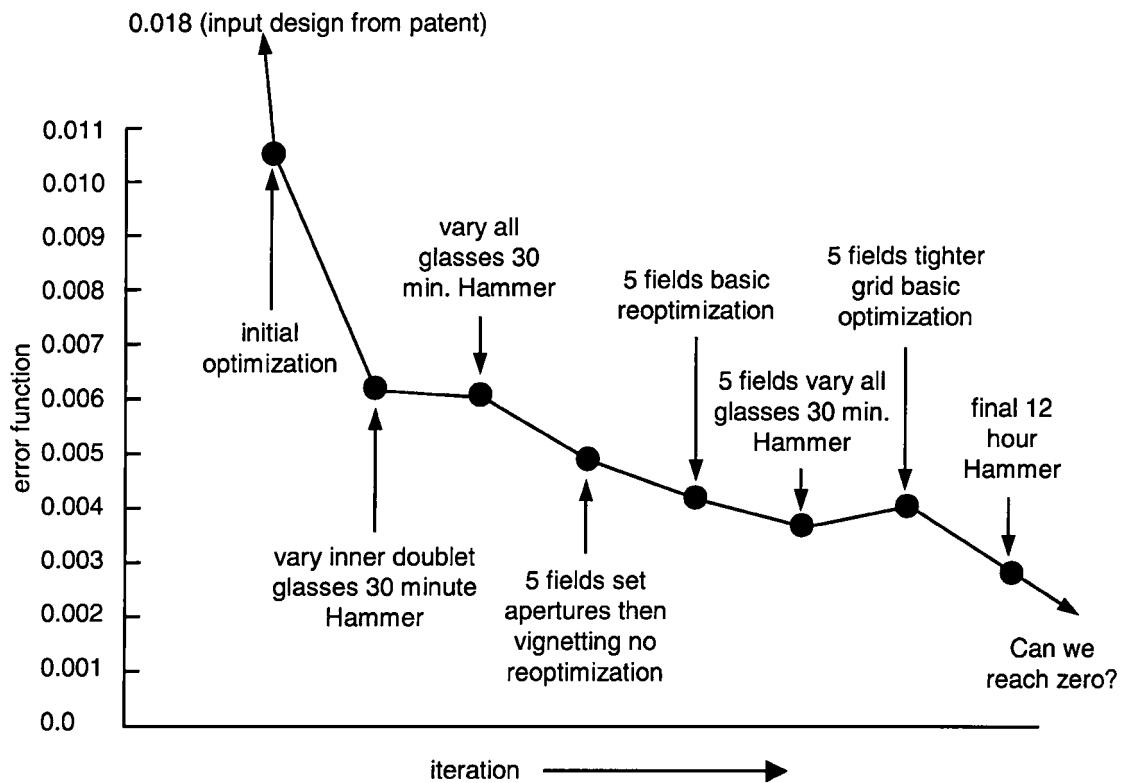
Figure 22.17
Pagi Diagram of
 $f/2.0$ Double Gauss



in the lens. We also can correlate steeper angles of incidence with greater aberration contributions (and more sensitive tolerances).

In the previous optimization sequence we took eight separate and independent steps in the optimization of the double Gauss lens. This is summarized graphically in Fig. 22.18, where we plot the steps taken in the abscissa and the error function in the ordinate. At the conclusion of each one of the individual steps, we reached a local minimum in the error function and we had to apply an outside influence prior to taking the next step. Notice that we raise the question “can we reach zero” in the error function? In order to reduce the error function more, we would need to add additional elements in order to further minimize the residual aberrations. A good way to think about the answer to this question is to recall the microlithography lens of Glatzel in Chap. 5. This lens has about seventeen elements, and while its performance is not perfect, it is clearly diffraction limited. Of course the specifications are not the same, but we can conclude that a lens with a similar level of complexity to the Glatzel lens may be required to bring the error function much closer to zero.

We now will stop the lens down to $f/4.0$ as in Fig. 22.19. It is important to evaluate the lens over the functional range over which it will be used, and in 35-mm photography, lenses are often used stopped down in f/number (at higher $f/\text{numbers}$).

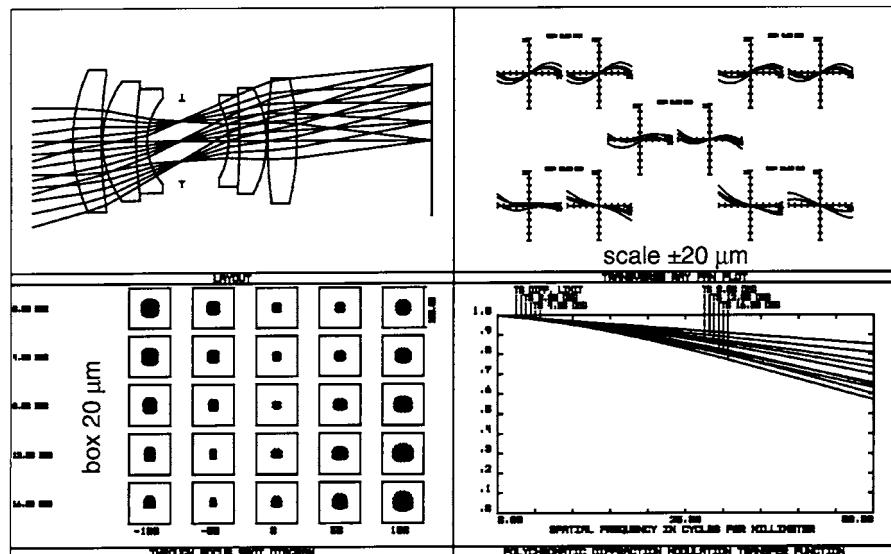
**Figure 22.18**

Progression of Error Function (Image Quality Portion Only) During Double Gauss Lens Design Example

We will now proceed to tolerance the lens. Figure 22.20 shows the Zemax input table where initial tolerances are input. Due to the reasonably tight level of the tolerances expected for our low f /number double Gauss lens, we will select a starting mix of tolerances which are representative of somewhat tight, yet achievable values. The following are the tolerances and the rationale:

- We will assume that prior to manufacture, all radii will be matched to existing testplates, and thus for the surface radii we will input power fit to testplate as four fringes. If we had to custom manufacture one or more testplates, then a specific radius tolerance would be necessary, indicating the accuracy to which the testplate were manufactured. As noted earlier, fitting 100% of the radii of a given lens design to existing testplates is generally done.

Figure 22.19
Final Design Stopped
Down to $f/4$



- All element thicknesses and airspaces are assumed to be ± 0.05 mm, a reasonable assumption.
- We have elected not to use surface decentrations as mathematically element wedge takes this into account.
- Surface tilts of 0.025-mm total indicator runout (TIR) on each surface is assumed for the element wedge. We have intentionally split the wedge between each surface of each element, so in effect the net total wedge of any given element is 0.05 mm TIR. We may be able to further refine the accuracy of our wedge model once we know the specific manufacturing methods.
- Surface irregularity is assumed in Zemax to be a mix of spherical aberration due to a fourth-order OPD contribution and astigmatism from cylinder. We have assumed one fringe of irregularity per surface.
- For refractive index and Abbe number we are using the default values of ± 0.0002 and ± 0.01 , respectively.
- The previous specifications are for surfaces. For elements we have decentration in x and y of ± 0.05 mm.
- For element tilt we are using 0.114° , which equates to approximately 0.05 mm TIR.

Default Tolerances

Surface Tolerances		Element Tolerances	
<input checked="" type="checkbox"/> Radius	<input type="radio"/> Millimeters: 0.200000	<input checked="" type="checkbox"/> Decenter X	0.050000
	<input checked="" type="radio"/> Fringes: 4.000000	<input checked="" type="checkbox"/> Decenter Y	0.050000
<input checked="" type="checkbox"/> Thickness	Millimeters: 0.050000	<input checked="" type="checkbox"/> Tilt X	Degrees: 0.114000
<input type="checkbox"/> Decenter X	Millimeters: 0.200000	<input checked="" type="checkbox"/> Tilt Y	Degrees: 0.114000
<input type="checkbox"/> Decenter Y	Millimeters: 0.200000		
<input checked="" type="checkbox"/> Tilt (TIR) X	<input type="radio"/> Millimeters: 0.025000		
	<input type="radio"/> Degrees: 0.200000		
<input checked="" type="checkbox"/> Tilt (TIR) Y	<input type="radio"/> Millimeters: 0.025000		
	<input type="radio"/> Degrees: 0.200000		
<input checked="" type="checkbox"/> S + A Irreg	Fringes: 1.000000		
<input type="checkbox"/> Zern Irreg	Fringes: 0.200000		
<input checked="" type="checkbox"/> Index	0.000200		
<input checked="" type="checkbox"/> Abbe	0.010000		
Start At Row:	1	<input checked="" type="checkbox"/> Use Focus Comp	
		<input type="button" value="OK"/>	<input type="button" value="Cancel"/>
		<input type="button" value="Save"/>	<input type="button" value="Load"/>
		<input type="button" value="Reset"/>	<input type="button" value="Help"/>

Figure 22.20
Input of Tolerance Values

Note at the bottom that we are using focus compensation as a compensator. What this means is that for each and every tolerance perturbation we will assume that the lens can be refocused. Since the lens is definitely refocused during final testing and assembly, this is a fair assumption.

We now show the initial output page, where the various assumptions are listed for the analysis. Note that the so-called merit function is the average of sagittal and tangential target orientation MTF at 30 line pairs/mm. Averaged over the field of view the overall nominal MTF is 0.836.

The results are shown as follows:

```

Analysis of Tolerances
Title: Final Final 12 Hour Hammer
Units are Millimeters.
Fast tolerancing mode is on. In this mode, all
compensators are ignored, except back focus error.

Merit: Diffraction MTF average S&T at 30.0000 1p/mm
Nominal Merit Function (MF) is 0.83564909
Test wavelength: 0.6328

```

Fields: User Defined Angle in degrees							
#	X-Field	Y-Field	Weight	VDX	VDY	VCX	VCY
1	0.000E+000	0.000E+000	1.000E+000	0.000	0.000	0.000	0.000
2	0.000E+000	4.000E+000	1.000E+000	0.000	0.006	0.009	0.091
3	0.000E+000	8.000E+000	1.000E+000	0.000	0.008	0.047	0.205
4	0.000E+000	1.200E+001	1.000E+000	0.000	0.000	0.100	0.337
5	0.000E+000	1.600E+001	1.000E+000	0.000	-0.004	0.248	0.493

Fringes of Power Fit to Testplate

We show here the drop in MTF for each tolerance, both averaged over the field of view (labeled "All" under the field column) as well as at each specific field of view. The first tolerance listed is "TFRN" which is fringes of power fit to testplate. For brevity, we show only surfaces 2 through 6 which is the front half of the lens (surface 1 is a dummy surface forward of the lens). The greatest degradation here is on surface 6 which is the strong concave surface prior to the stop, and at fields 1 and 2 the MTF drops approximately 0.012. This is not a large MTF drop at all, and, in fact, most of the power fit to testplate tolerances can likely be increased from four fringes to five or more fringes with little effect.

Sensitivity Analysis:

Type	Sf1	Sf2	Field	Minimum			Maximum		
				Value	MF	Change	Value	MF	Change
TFRN	2	All	+4.000000	0.837422	0.001773	+4.000000	0.832193	-0.003456	
		1	0.806590	0.002522			0.798613	-0.005454	
		2	0.828502	0.002910			0.820480	-0.005112	
		3	0.854749	0.002028			0.849222	-0.003499	
		4	0.850717	0.001319			0.847165	-0.002233	
		5	0.851824	-0.000270			0.851948	-0.000145	
TFRN	3	All	-4.000000	0.833350	-0.002299	-4.000000	0.836925	0.001276	
		1	0.801921	-0.002147			0.805537	0.001469	
		2	0.823507	-0.002085			0.827025	0.001433	
		3	0.851510	-0.001211			0.853042	0.000321	
		4	0.847407	-0.001991			0.850012	0.000614	
		5	0.847876	-0.004217			0.854610	0.002517	

TFRN	4	All	-4.000000	0.837272	0.001623	4.000000	0.831326	-0.004323
		1		0.806835	0.002767		0.796489	-0.007579
		2		0.829831	0.004239		0.817930	-0.007662
		3		0.856489	0.003768		0.847042	-0.005679
		4		0.850653	0.001255		0.846977	-0.002421
		5		0.847577	-0.004516		0.855512	0.003419
TFRN	5	All	-4.000000	0.835808	0.000158	4.000000	0.835468	-0.000181
		1		0.804200	0.000132		0.803907	-0.000161
		2		0.825782	0.000190		0.825377	-0.000215
		3		0.852948	0.000226		0.852473	-0.000248
		4		0.849651	0.000253		0.849126	-0.000272
		5		0.852092	-0.000001		0.852078	-0.000015
TFRN	6	All	-4.000000	0.828345	-0.007304	4.000000	0.837541	0.001892
		1		0.792139	-0.011928		0.806558	0.002491
		2		0.813809	-0.011783		0.830302	0.004710
		3		0.844253	-0.008468		0.856661	0.003940
		4		0.845195	-0.004203		0.850895	0.001497
		5		0.854287	0.002194		0.848446	-0.003647

Thickness: Both Element Thicknesses and Airspaces

We show next thickness tolerances (TTHI) of ± 0.05 mm. The largest MTF drop is approximately 0.083 on axis for thicknesses 4 and 5, which is the inner doublet prior to the stop. This is of some significance, and we should revisit these tolerances after we complete the analysis. Note that the most sensitive tolerances are highlighted via an asterisk.

Type	Sf1	Sf2	Field	Minimum			Maximum		
				Value	MF	Change	Value	MF	Change
TTHI	2	3	All	-0.050000	0.835880	0.000231	0.050000	0.835116	-0.000534
		1		0.804020	-0.000048		0.804120	0.000053	
		2		0.825708	0.000116		0.825453	-0.000139	
		3		0.851833	-0.000888		0.853440	0.000718	
		4		0.848441	-0.000957		0.849847	0.000449	
		5		0.855198	0.003105		0.848101	-0.003992	
TTHI	3	6	All	-0.050000	0.835371	-0.000279	0.050000	0.829620	-0.006029
		1		0.806188	0.002121		0.792026	-0.012041	
		2		0.829775	0.004184		0.815661	-0.009931	
		3		0.858979	0.006258		0.844557	-0.008165	
		4		0.851557	0.002159		0.844201	-0.005197	
		5		0.835523	-0.016570		0.860443	0.008350	
TTHI	4	6	All	-0.050000	0.789088	-0.046561	0.050000	0.756450	-0.079199
		1		0.720801	-0.083266		0.660283	-0.143785*	
		2		0.783468	-0.042124		0.715721	-0.109871*	
		3		0.850413	-0.002309		0.788121	-0.064600	
		4		0.840267	-0.009130		0.816846	-0.032552	
		5		0.777086	-0.075007		0.851939	-0.000154	

TTHI	5	6	All	-0.050000	0.789928	-0.045721	0.050000	0.757571	-0.078078
		1			0.722150	-0.081918		0.661822	-0.142246*
		2			0.784430	-0.041162		0.717043	-0.108549*
		3			0.850703	-0.002018		0.789293	-0.063428
		4			0.840283	-0.009115		0.817884	-0.031514
		5			0.778232	-0.073861		0.852130	0.000037
TTHI	6	7	All	-0.050000	0.835714	0.000065	0.050000	0.835576	-0.000073
		1			0.804068	0.000000		0.804067	-0.000000
		2			0.825633	0.000041		0.825550	-0.000042
		3			0.852748	0.000027		0.852690	-0.000031
		4			0.849347	-0.000050		0.849433	0.000035
		5			0.852432	0.000339		0.851731	-0.000362

Element Decentration

We show next several element decentrations (TEDX and TEDY) as well as element tilts (TETX and TETY). Most of the MTF drops here are in the order of 0.02 to 0.05, and we should revisit these later.

Type	Sf1	Sf2	Field	Value	5		0.851583 -0.000510		0.852224 0.000131		
					Minimum	Maximum	Value	MF	Change	Value	MF
TEDX	2	3	All	-0.050000	0.817067	-0.018582	0.050000	0.817067	-0.018582	0.817067	-0.018582
		1			0.781078	-0.022990		0.781078	-0.022990		
		2			0.803667	-0.021925		0.803667	-0.021925		
		3			0.832935	-0.019786		0.832935	-0.019786		
		4			0.832676	-0.016722		0.832676	-0.016722		
		5			0.842073	-0.010020		0.842073	-0.010020		
TEDY	2	3	All	-0.050000	0.808191	-0.027458	0.050000	0.817327	-0.018322	0.817327	-0.018322
		1			0.781078	-0.022990		0.781078	-0.022990		
		2			0.795995	-0.029596		0.814195	-0.011397		
		3			0.835070	-0.017651		0.823512	-0.029209		
		4			0.826343	-0.023055		0.821939	-0.027459		
		5			0.807515	-0.044578		0.853223	0.001130		
TETX	2	3	All	-0.114000	0.809352	-0.026297	0.114000	0.808853	-0.026796	0.808853	-0.026796
		1			0.790527	-0.013541		0.790527	-0.013541		
		2			0.812954	-0.012638		0.802454	-0.023138		
		3			0.805097	-0.047624		0.838547	-0.014174		
		4			0.798082	-0.051315		0.824232	-0.025166		
		5			0.844726	-0.007367		0.793068	-0.059025		
TETY	2	3	All	-0.114000	0.813819	-0.021830	0.114000	0.813819	-0.021830	0.813819	-0.021830
		1			0.790527	-0.013541		0.790527	-0.013541		
		2			0.809240	-0.016352		0.809240	-0.016352		
		3			0.829636	-0.023086		0.829636	-0.023086		
		4			0.819439	-0.029959		0.819439	-0.029959		
		5			0.822740	-0.029354		0.822740	-0.029354		
TEDX	4	6	All	-0.050000	0.813285	-0.022364	0.050000	0.813285	-0.022364	0.813285	-0.022364
		1			0.783273	-0.020795		0.783273	-0.020795		
		2			0.803831	-0.021761		0.803831	-0.021761		
		3			0.828743	-0.023978		0.828743	-0.023978		
		4			0.823886	-0.025512		0.823886	-0.025512		
		5			0.831138	-0.020955		0.831138	-0.020955		
TEDY	4	6	All	-0.050000	0.801162	-0.034487	0.050000	0.788111	-0.047538	0.788111	-0.047538

				1	0.783273 -0.020795	0.783273 -0.020795
				2	0.805776 -0.019816	0.790366 -0.035225
				3	0.796606 -0.056116	0.815828 -0.036893
				4	0.790714 -0.058684	0.790983 -0.058415
				5	0.833211 -0.018882	0.763443 -0.088651
TETX	4	6	All	-0.114000	0.785336 -0.050313	0.114000 0.801904 -0.033745
				1	0.789832 -0.014236	0.789832 -0.014236
				2	0.793959 -0.031633	0.808478 -0.017114
				3	0.811466 -0.041255	0.797176 -0.055545
				4	0.779829 -0.069569	0.789780 -0.059618
				5	0.755544 -0.096549	0.826712 -0.025381
TETY	4	6	All	-0.114000	0.818623 -0.017027	0.114000 0.818623 -0.017027
				1	0.789832 -0.014236	0.789832 -0.014236
				2	0.810120 -0.015472	0.810120 -0.015472
				3	0.834414 -0.018307	0.834414 -0.018307
				4	0.828629 -0.020769	0.828629 -0.020769
				5	0.834238 -0.017856	0.834238 -0.017856

Element Wedge

We show the tolerances for element wedge as the total indicator runout (TIRX and TIRY). We had assigned 0.025 mm for each surface, which is, in effect, 0.05 mm for the element. Most of the surfaces listed are not too sensitive, except for surface 4 which is the front of the forward doublet where the largest MTF drop is approximately 0.07 at one of the outer field positions.

Type	Sf1	Sf2	Field	Minimum			Maximum		
				Value	MF	Change	Value	MF	Change
TIRX	2	All	-0.025000	0.826102 -0.009547	0.025000	0.826102 -0.009547	0.791277 -0.012791	0.813671 -0.011921	0.842599 -0.010122
				1	0.791277 -0.012791				
				2	0.813671 -0.011921				
				3	0.842599 -0.010122				
				4	0.841593 -0.007805				
				5	0.848177 -0.003917				
TIRY	2	All	-0.025000	0.828801 -0.006848	0.025000	0.820179 -0.015470	0.791277 -0.012791	0.807437 -0.018155	0.842928 -0.009793
				1	0.791277 -0.012791				
				2	0.822472 -0.003120				
				3	0.840980 -0.011741				
				4	0.839645 -0.009753				
				5	0.856950 0.004856				
TIRX	3	All	-0.025000	0.830928 -0.004721	0.025000	0.830928 -0.004721	0.800116 -0.003952	0.821408 -0.004184	0.847847 -0.004875
				1	0.800116 -0.003952				
				2	0.821408 -0.004184				
				3	0.847847 -0.004875				
				4	0.843688 -0.005709				
				5	0.846723 -0.005370				
TIRY	3	All	-0.025000	0.827219 -0.008430	0.025000	0.836174 0.000524	0.800116 -0.003952	0.827514 0.001922	0.856938 0.004216
				1	0.800116 -0.003952				
				2	0.817082 -0.008510				
				3	0.841542 -0.011179				
				4	0.836525 -0.012873				
				5	0.845017 -0.007076				

TIRX	4	All	-0.025000	0.830680	-0.004969	0.025000	0.830680	-0.004969
		1		0.800850	-0.003217		0.800850	-0.003217
		2		0.821652	-0.003940		0.821652	-0.003940
		3		0.847230	-0.005491		0.847230	-0.005491
		4		0.842622	-0.006776		0.842622	-0.006776
		5		0.845819	-0.006274		0.845819	-0.006274
TIRY	4	All	-0.025000	0.794785	-0.040864	0.025000	0.765899	-0.069750
		1		0.800850	-0.003217		0.800850	-0.003217
		2		0.806401	-0.019191		0.802323	-0.023269
		3		0.789915	-0.062806		0.795301	-0.057420
		4		0.775756	-0.073642		0.736589	-0.112809*
		5		0.802509	-0.049584		0.710186	-0.141907*
TIRX	5	All	-0.025000	0.835219	-0.000430	0.025000	0.835219	-0.000430
		1		0.803645	-0.000423		0.803645	-0.000423
		2		0.825162	-0.000430		0.825162	-0.000430
		3		0.852279	-0.000443		0.852279	-0.000443
		4		0.848959	-0.000439		0.848959	-0.000439
		5		0.851658	-0.000435		0.851658	-0.000435
TIRY	5	All	-0.025000	0.834858	-0.000791	0.025000	0.835580	-0.000069
		1		0.803645	-0.000423		0.803645	-0.000423
		2		0.824740	-0.000852		0.825634	0.000042
		3		0.851027	-0.001694		0.853591	0.000870
		4		0.847527	-0.001871		0.850388	0.000990
		5		0.852874	0.000780		0.850362	-0.001731
TIRX	6	All	-0.025000	0.820268	-0.015381	0.025000	0.820268	-0.015381
		1		0.791865	-0.012202		0.791865	-0.012202
		2		0.811931	-0.013661		0.811931	-0.013661
		3		0.835974	-0.016747		0.835974	-0.016747
		4		0.830315	-0.019082		0.830315	-0.019082
		5		0.835288	-0.016805		0.835288	-0.016805
TIRY	6	All	-0.025000	0.734318	-0.101331	0.025000	0.765926	-0.069723
		1		0.791865	-0.012202		0.791865	-0.012202
		2		0.784413	-0.041179		0.791466	-0.034126
		3		0.767164	-0.085557		0.752842	-0.099880
		4		0.695016	-0.154382		0.730450	-0.118948
		5		0.659546	-0.192547		0.768909	-0.083184

Surface Irregularity

We show the sensitivities for surface irregularity (TIRR) where one fringe is assumed. Most of the sensitivities are reasonable, and we may be able to loosen some of these tolerances to perhaps two to three fringes of irregularity.

Type	Sf1	Sf2	Field	Minimum			Maximum		
				Value	MF	Change	Value	MF	Change
TIRR	2	All	-1.000000	0.833455	-0.002194	1.000000	0.828633	-0.007017	
		1		0.803273	-0.000794		0.792029	-0.012039	
		2		0.822489	-0.003103		0.818111	-0.007481	
		3		0.845813	-0.006908		0.850890	-0.001831	
		4		0.845717	-0.003680		0.845692	-0.003706	
		5		0.855439	0.003345		0.843624	-0.008470	

TIRR	3	All	-1.000000	0.826018	-0.009632	1.000000	0.832446	-0.003203
		1		0.788850	-0.015218		0.802526	-0.001542
		2		0.815685	-0.009907		0.821513	-0.004079
		3		0.849174	-0.003547		0.844043	-0.008678
		4		0.843523	-0.005875		0.844021	-0.005377
		5		0.840126	-0.011967		0.855549	0.003456
TIRR	4	All	-1.000000	0.833276	-0.002373	1.000000	0.828988	-0.006661
		1		0.802227	-0.001840		0.791542	-0.012526
		2		0.822124	-0.003468		0.817340	-0.008252
		3		0.846462	-0.006259		0.850571	-0.002151
		4		0.846556	-0.002842		0.846653	-0.002745
		5		0.854702	0.002609		0.846504	-0.005589
TIRR	5	All	-1.000000	0.835714	0.000065	1.000000	0.835579	-0.000070
		1		0.804195	0.000127		0.803933	-0.000134
		2		0.825669	0.000077		0.825509	-0.000083
		3		0.852715	-0.000006		0.852723	0.000001
		4		0.849428	0.000030		0.849364	-0.000034
		5		0.852171	0.000077		0.852014	-0.000080
TIRR	6	All	-1.000000	0.828973	-0.006676	1.000000	0.833132	-0.002517
		1		0.791582	-0.012486		0.801799	-0.002269
		2		0.816892	-0.008700		0.822012	-0.003580
		3		0.850059	-0.002662		0.846626	-0.006095
		4		0.846769	-0.002629		0.846613	-0.002784
		5		0.847286	-0.004807		0.854362	0.002269

Refractive Index and Abbe Number

Here we show the sensitivities to refractive index and Abbe number (TIND and TABB). These are quite insensitive and could be loosened if there is a reason to do so.

Type	Sf1	Sf2	Field	Minimum			Maximum		
				Value	MF	Change	Value	MF	Change
TIND	2	All	-0.000200	0.836125	0.000476	0.000200	0.835101	-0.000548	
		1		0.804687	0.000619		0.803357	-0.000711	
		2		0.826309	0.000718		0.824797	-0.000795	
		3		0.853150	0.000429		0.852222	-0.000500	
		4		0.849676	0.000279		0.849054	-0.000344	
		5		0.852362	0.000269		0.851773	-0.000320	
TIND	4	All	-0.000200	0.836882	0.001233	0.000200	0.833834	-0.001815	
		1		0.805930	0.001862		0.801194	-0.002874	
		2		0.827573	0.001981		0.822850	-0.002742	
		3		0.853965	0.001244		0.850964	-0.001758	
		4		0.850101	0.000703		0.848374	-0.001024	
		5		0.852190	0.000097		0.851854	-0.000239	
TIND	5	All	-0.000200	0.833863	-0.001786	0.000200	0.836876	0.001227	
		1		0.801434	-0.002633		0.805796	0.001728	
		2		0.823050	-0.002542		0.827435	0.001844	
		3		0.851014	-0.001708		0.853923	0.001201	
		4		0.848190	-0.001208		0.850245	0.000847	
		5		0.851592	-0.000501		0.852392	0.000299	
TIND	8	All	-0.000200	0.834542	-0.001107	0.000200	0.836503	0.000853	
		1		0.802123	-0.001945		0.805510	0.001442	
		2		0.823845	-0.001747		0.826991	0.001399	

				3	0.851480	-0.001242		0.853779	0.001058
				4	0.848730	-0.000668		0.849976	0.000578
				5	0.852504	0.000411		0.851640	-0.000454
TIND	9	All	-0.000200	0.836471	0.000822	0.000200	0.834587	-0.001062	
				1	0.805510	0.001442		0.802141	-0.001927
				2	0.827001	0.001409		0.823849	-0.001743
				3	0.853712	0.000991		0.851557	-0.001164
				4	0.849825	0.000427		0.848893	-0.000505
				5	0.851674	-0.000419		0.852484	0.000391
TIND	11	All	-0.000200	0.836082	0.000433	0.000200	0.835162	-0.000487	
				1	0.804780	0.000712		0.803254	-0.000814
				2	0.826269	0.000677		0.824842	-0.000750
				3	0.853184	0.000463		0.852216	-0.000505
				4	0.849615	0.000218		0.849156	-0.000241
				5	0.852076	-0.000018		0.852099	0.000006
TABB	2	All	-0.010000	0.835612	-0.000037	0.010000	0.835684	0.000035	
				1	0.804060	-0.000007		0.804072	0.000005
				2	0.825548	-0.000044		0.825633	0.000041
				3	0.852631	-0.000091		0.852810	0.000088
				4	0.849292	-0.000106		0.849501	0.000103
				5	0.852145	0.000051		0.852039	-0.000054
TABB	4	All	-0.010000	0.835571	-0.000079	0.010000	0.835717	0.000068	
				1	0.804047	-0.000020		0.804075	0.000008
				2	0.825499	-0.000093		0.825674	0.000082
				3	0.852548	-0.000174		0.852885	0.000164
				4	0.849205	-0.000193		0.849581	0.000183
				5	0.852163	0.000070		0.852015	-0.000078
TABB	5	All	-0.010000	0.835784	0.000135	0.010000	0.835460	-0.000189	
				1	0.804059	-0.000009		0.804007	-0.000060
				2	0.825751	0.000159		0.825371	-0.000221
				3	0.853054	0.000332		0.852336	-0.000386
				4	0.849772	0.000374		0.848978	-0.000420
				5	0.851954	-0.000139		0.852194	0.000101

We will now list the 10 worst offenders, in other words, the parameters and their associated tolerances giving the biggest drop in MTF. These data are for the average MTF computed over the entire field of view. Note that most of the more sensitive tolerances are element wedges. We also see here the nominal MTF (averaged over the field) to be 0.84, the estimated change in MTF of -0.23, and the estimated MTF of 0.61. There is additional statistical information including the expected compensator amount.

Worst offenders:						
Type	Sf1	Sf2	Value	MF	Change	
TIRY	6	-0.025000	0.734318	-0.101331		
TTHI	4	6	0.050000	0.756450	-0.079199	
TTHI	5	6	0.050000	0.757571	-0.078078	
TIRY	4	0.025000	0.765899	-0.069750		
TIRY	6	0.025000	0.765926	-0.069723		
TIRY	8	-0.025000	0.778044	-0.057605		
TIRX	8	-0.025000	0.778506	-0.057143		
TIRX	8	0.025000	0.778506	-0.057143		
TIRY	8	0.025000	0.779299	-0.056350		
TETX	4	6	-0.114000	0.785336	-0.050313	

The net prediction for the expected MTF at 30 line pairs/mm is shown here. This is for the average MTF over the field of view and we see that the prediction is for an MTF of 0.64 at 30 line pairs/mm. Note also that the expected total range in back focus adjustment is approximately $\pm 180 \mu\text{m}$, with a standard deviation of 31 μm .

```

Nominal MTF      :      0.84
Estimated change :     -0.19
Estimated MTF    :      0.64
Merit Statistics :
  Mean           :      0.824034
  Standard Deviation : 0.018316
Compensator Statistics:
Change in back focus:
  Minimum        :     -0.183763
  Maximum        :      0.184228
  Mean           :      0.000007
  Standard Deviation : 0.031307

```

Perhaps the most important way to assess the overall tolerance situation is via a *Monte Carlo analysis*. In this analysis, every parameter is perturbed between its plus and minus tolerance extremes according to a normal probability distribution. The MTF is then computer averaged over the field of view and at each of the five separate fields. The resulting statistics shows for the 20 Monte Carlo samples the nominal, best, worst, mean, and standard deviation in the MTF. Recall that our MTF goal is 0.5 at 30 line pairs/mm. Field 4 (75% of the way to the corner) shows a worst MTF encountered of 0.494. The best MTF at field 4 was 0.821. Here we show the results of 20 Monte Carlo samples.

Monte Carlo Analysis:

Number of trials: 20

Statistics: Normal Distribution

				0.0, 0.0 0.0, 0.3 0.0, 0.5 0.0, 0.8 0.0, 1.0									
Trial	Merit	Change	Field 1	Field 2	Field 3	Field 4	Field 5						
1	0.671008	-0.164641	0.598434	0.598588	0.674171	0.743139	0.784013						
2	0.739904	-0.095745	0.654500	0.751300	0.815063	0.781986	0.725598						
3	0.683551	-0.152098	0.636928	0.662700	0.710189	0.703255	0.711801						
4	0.671622	-0.164027	0.586082	0.620614	0.682513	0.719141	0.789717						
5	0.752621	-0.083028	0.727852	0.714027	0.743741	0.778635	0.811677						
6	0.815043	-0.126764	0.663265	0.709160	0.728300	0.712060	0.737265						
7	0.708885	-0.108246	0.678229	0.718370	0.741885	0.738249	0.768576						
8	0.727403	-0.092848	0.729252	0.767542	0.745248	0.719039	0.755914						
9	0.742801	-0.108210	0.654835	0.687834	0.761382	0.784023	0.773559						
10	0.678215	-0.157435	0.603121	0.639159	0.684315	0.718385	0.774066						
11	0.691634	-0.144015	0.654557	0.666323	0.719306	0.721229	0.702888						
12	0.691634	-0.094413	0.688700	0.713367	0.762045	0.765506	0.789961						
13	0.741237	-0.091746	0.647524	0.705033	0.773729	0.700739	0.840690						
14	0.641301	-0.194348	0.634102	0.701845	0.719443	0.628549	0.548493						

```

15 0.653456 -0.182194 0.591334 0.604764 0.638324 0.687568 0.779041
16 0.805344 -0.030305 0.766817 0.809427 0.839262 0.821381 0.797463
17 0.754233 -0.081416 0.776285 0.766568 0.766268 0.737186 0.728399
18 0.681817 -0.153832 0.600345 0.637742 0.701103 0.731612 0.767892
19 0.764597 -0.071052 0.734169 0.747044 0.781917 0.777415 0.787125
20 0.558923 -0.276726 0.570030 0.628170 0.579650 0.494046 0.534242

Nominal 0.835649          0.804068 0.825592 0.852721 0.849398 0.852093
Best    0.805344          0.776285 0.809427 0.839262 0.821381 0.840690
Worst   0.558923          0.570030 0.598588 0.579650 0.494046 0.534242
Mean    0.706995          0.659818 0.692479 0.728393 0.728107 0.745419
Std Dev 0.053359          0.059708 0.057695 0.058114 0.068625 0.075600

Compensator Statistics:
Change in back focus:
Minimum      : -0.159362
Maximum      : 0.277543
Mean         : -0.031688
Standard Deviation : 0.116491

```

The final results of the Monte Carlo analysis are shown next, where we see that 90% of the lenses have an MTF of greater than or equal to 0.703. The standard deviation in the compensator motion (refocusing) is 0.116 mm.

```

90% of Monte Carlo lenses have an MTF above 0.703
50% of Monte Carlo lenses have an MTF above 0.774
10% of Monte Carlo lenses have an MTF above 0.812

```

The final result of the tolerance analysis is that the lens will, with the input tolerances, meet our MTF performance goal of 0.5 minimum at 30 line pairs/mm. If we were to take the analysis further, we would discuss the more sensitive parameters with the optical shop to see if we can tighten them. Then we would tighten these tolerances and simultaneously loosen many of the less sensitive tolerances so that the net result is meeting the MTF performance requirement while making the lens more producible at a lower cost and higher confidence.

The previous tolerances were computered for one side of the field of view. Due to the asymmetrical nature of some tolerances, both sides of the field should be modeled. When this is done, the 90% confidence level MTF at 30 line pairs/mm reduces to 0.634, still well above our goal of 0.5.

Digital Camera Lens

This case study is based on a VGA digital camera, which has some rather unique specifications. Specifically, some time ago we bought a digital camera which has a $\frac{1}{3}$ -in CCD chip. The lens is $f/2.0$, and the camera

TABLE 22.3

Digital Camera
Lens Design
Example

Parameter	Specification
Sensor type	CCD
Sensor size	1/3 in (3.6 × 4.8 mm, 6-mm diagonal)
Number of pixels	640 × 480
Pixel pitch	7.5 μm
Lens <i>f</i> /number	<i>f</i> /2.0
Lens focal length	4.8 mm
Comparable focal length in 35-mm camera	35-mm focal length
Stated depth of field	533-mm to infinity

manual states that objects from 533 mm (21 in) to infinity will be in focus. Many of us who have worked a lot in 35-mm camera photography will remember that if we focus on the front of someone's nose at *f*/2.0, that person's earlobes will be out of focus. The depth of field is incredibly small at such a low *f*/number. What is it that allows for such a large depth of field in our digital camera? The answer will be given later. First, we will summarize in Table 22.3 the specifications for our camera lens.

The Airy disk diameter is 2.8 μm, or about one-third of a pixel. Let us use Newton's equation, which relates object distance to the amount of defocus. Newton's equation states that

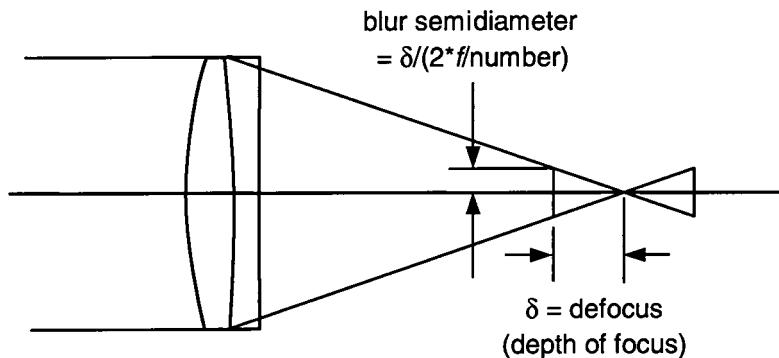
$$-xx' = (\text{focal length})^2 \quad x' = \frac{f^2}{-x} = \frac{f^2}{\text{object distance}} = \text{defocus}$$

where x' is the amount of refocus required for an object at distance x . If we compute the defocus required for a given object distance, we can easily determine the blur diameter by multiplying the defocus by $1/(f/\text{number}) = 0.5$. Figure 22.21 shows the situation.

Table 22.4 shows the defocus along with the associated blur diameter.

What this means is that if we have an otherwise perfect lens at *f*/2.0 focused for an object at infinity, the image distance will change by the amount "δ image distance" as a function of the object distance. If our sensor were to remain fixed at the infinity focus position, then the image would blur to the diameter indicated in the third column. Thus, an object at 0.5 m will blur to a diameter of 23 μm. This is approximately three pixels of image blur, which seems excessive. However, what if we

Figure 22.21
Computing Blur
Diameter for Depth-
of-Focus Analysis



select an intermediate object distance at which to focus our lens nominally so that the blur is equalized at infinity and at 0.5 m. This distance is approximately 1 m, and the residual image blur with the object at infinity and at 0.5 m is approximately 11 μm which is in the order of 1.5 pixels. Thus, the bottom line is that at the factory the lens will be focused for an object distance between 1 and 2 m, and in use the maximum image blur diameter from a point object everywhere from 0.5 m to infinity will be about 10 μm , or about 1.5 pixels. This large depth of field achieved with an $f/2.0$ lens explains how a digital camera can have a fixed focus.

Why is it then that the $f/2$ 35-mm camera lens cannot have a fixed focus, and even with the adjustable focus, some objects in the field of view are less sharp than the others? To answer this question, we will compare two lenses: a digital camera lens and a 35-mm camera lens. Let us choose the 35-mm camera lens to have the same field of view as the

TABLE 22.4 Depth-of-Focus Calculation for $f/2$ Digital Camera Lens

Object Distance	δ Image Distance (μm) (Sensor Focused for Infinity)	Blur Diameter (μm) (Sensor Focused for Infinity)	δ Image Distance (μm) (Sensor Focused for 1-m Distance)	Blur Diameter (μm) (Sensor Focused for 1-m Distance)
Infinity	0	0	-23	11.5
3	7.68	3.84	-15.3	7.7
2	11.5	5.75	-11.5	5.75
1	23.0	11.54	0	0
0.5	46.1	23	23	11.5

digital camera. This determines the focal length of the camera lens to be 35 mm. If both cameras are focused to infinity, we can compare the depth of field allowing the same angular blur in both cameras. This can be expressed as the equal relative linear blur in the image plane:

$$\left[\frac{\text{Linear blur}}{\text{Linear field of view}} \right]_{\text{35-mm camera}} = \left[\frac{\text{Linear blur}}{\text{Linear field of view}} \right]_{\text{digital camera}}$$

$$\frac{f_1^2}{D_1 f\# \text{fov}_1} = \frac{f_2^2}{D_2 f\# \text{fov}_2}$$

where f_1 and f_2 are, respectively, focal lengths of two cameras, D_1 and D_2 are distances at which the angular image blur is acceptable and equal in both cases, and fov_1 and fov_2 are the respective image sizes.

$$D_1 = D_2 \left(\frac{f_1}{f_2} \right)^2 \frac{\text{fov}_2}{\text{fov}_1} = D_2 \left(\frac{f_1}{f_2} \right)^2 \frac{f_2}{f_1} = D_2 \frac{f_1}{f_2}$$

$$D_1 = 1500 \frac{35}{4.8} = 10937$$

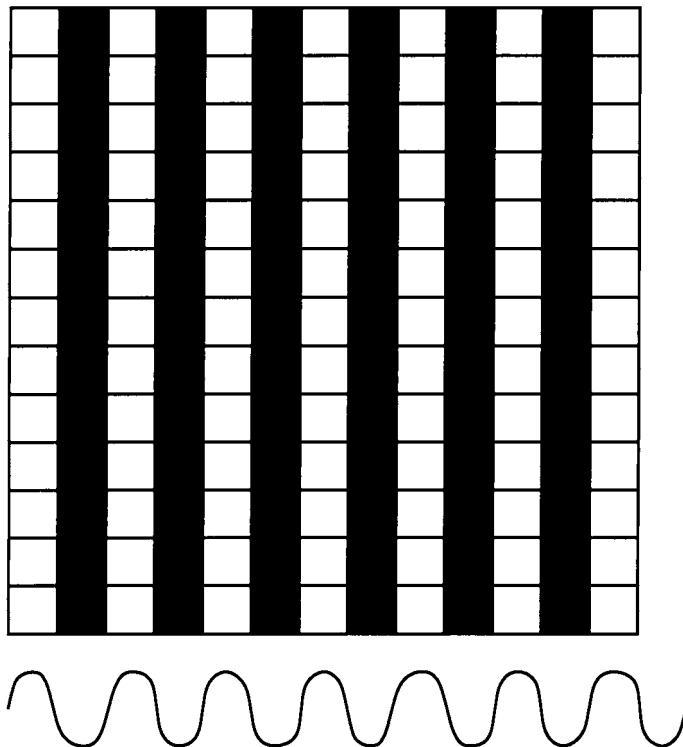
We see that if with the digital camera we can go from infinity to 1.5 m, for the same allowable image blur, with the 35-mm camera lens we can go from infinity to only about 11 m. Depth of field is inversely proportional to the focal length of the lens.

Prior to designing our lenses, we need to determine at what spatial frequency the lens should be evaluated. Consider Fig. 22.22, where we show the representation of a pixelated sensor such as a CCD. For our VGA CCD sensor the pixel pitch is 7.5 μm . The maximum spatial frequency of an image, which can effectively be resolved by a pixelated sensor without aliasing, is the spatial frequency where the bright and dark bars line up with adjacent rows or columns of the sensor as shown in Fig. 22.22. This frequency is called the *Nyquist frequency*. At higher-image spatial frequencies we will get so-called aliasing, where the image is undersampled by the sensor. Angled lines look like staircases due to the undersampling. For our camera case study the 7.5- μm pixel pitch results in a Nyquist frequency of 66.6 line pairs/mm. We will thus evaluate our lens performance at this value.

In order to determine the quality of the lens that we bought, we decided to measure two basic characteristics: camera resolution and distortion. This may be useful and serve as a reference during the design of the lens. To determine the resolution, we took the picture of a resolution chart, shown in Fig. 22.23 from a distance of 1 m. We measured the smallest diameter in the photo of the chart where the radial lines were resolved,

Figure 22.22

Imagery onto
Pixelated Sensor
and the Nyquist
Frequency



and found a corresponding width of a line pair, which was 52 line pairs/mm at the CCD chip. This corresponds to a modulation of only a few percent. Taking into account normal manufacturing errors, we can conclude to a close approximation that the nominal design of this camera lens has an MTF in the order of 0.3 to 0.4 at 52 line pairs/mm. Distortion was measured from the picture of an object with a straight edge whose geometry is shown in Fig. 22.24. From the measured sag of the bowed image of the straight line we calculated the distortion to be less than 3%.

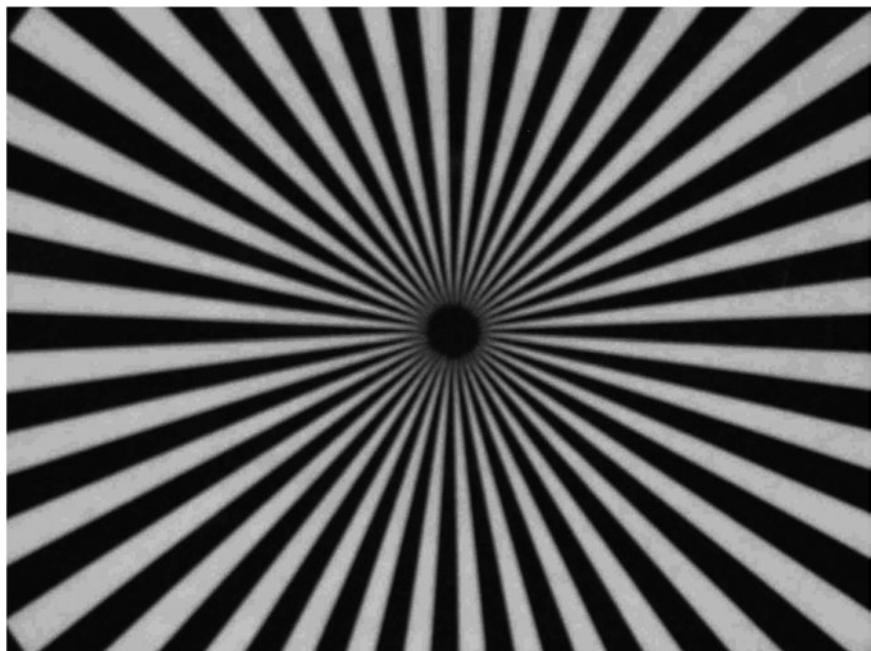
Note that the previous assessment of image quality and distortion was done very quickly with extremely rudimentary equipment and without removing the lens from the camera. These forms of tests can often be extremely useful, even though they are not highly quantitative.

We will now proceed to look at several candidate design solutions for the lens. The design parameters are as follows:

- Focal length of 4.8 mm
- $f/2$ lens

Figure 22.23

Image of Resolution Chart Taken with Digital Camera

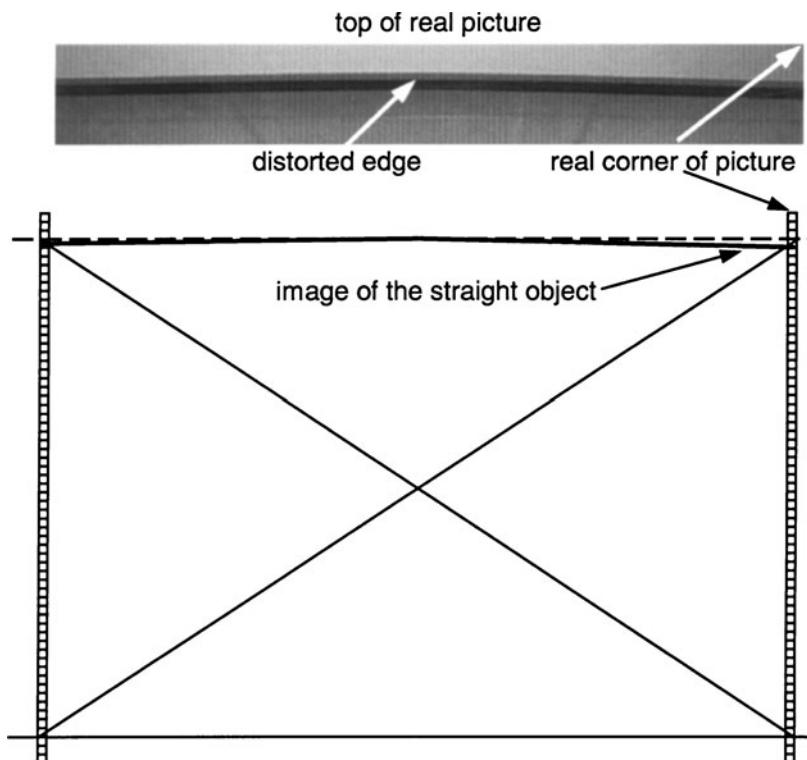


- Field of view (diagonal) of 64°
- Nyquist frequency of 66 line pairs/mm

Design with Three Glass Elements

Figure 22.25 shows a design for an all-glass lens with spherical surfaces. The field of view is quite large, and we would not expect to have a nicely behaved lens with small angles of incidence as the rays proceed through the lens. In this lens, rays enter the second component at very large incident angles, and also reach the detector at very large angles at the corner of the CCD. After optimizing different configurations, one criterion that is used to determine which lens is better than the other is how strongly rays refract on each surface throughout the lens. The manufacturing tolerances have to be tighter in the locations of strong ray bending, so that the lens with the smoother ray travel through the lens is preferable.

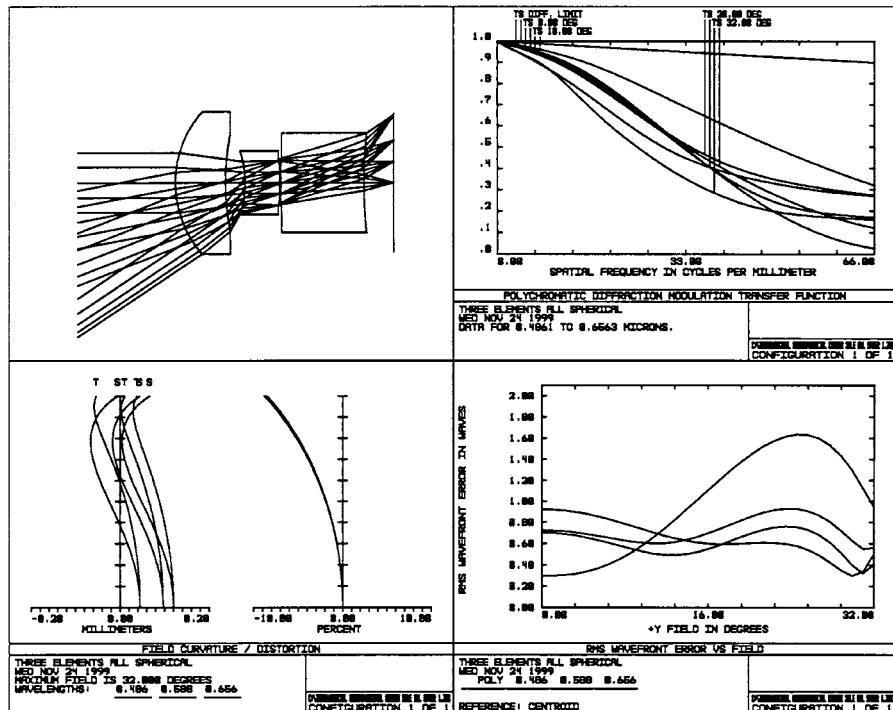
Figure 22.24
Measurement of
Lens Distortion



We will analyze a few configurations comparing their performance shown in four diagrams. The first is the lens layout. The second is the MTF curve for four field angles shown to the Nyquist frequency of 66 line pairs/mm. The third diagram is the field curvature and the distortion curve, given on the scale of 10%. The fourth diagram is the rms wavefront error plotted on a two-wave scale, as a function of field of view. The MTF of this lens should be higher, although the measured resolution of the digital camera suggests that its MTF is probably a little lower than the MTF of our design shown here. The distortion is definitely unacceptable, and it will have to be more tightly controlled. In the next step we increased the weights on the chief ray heights in the merit function, in order to reduce the distortion below 3%. We also allowed both surfaces of the third component to be aspheric. The result was a reduced distortion down to 4%, but the MTF and the lens shape remained the same.

Figure 22.25

Three All-Spherical-Glass Elements for Digital Camera Lens



Design with Four Glass Elements, Two of Them Aspheric

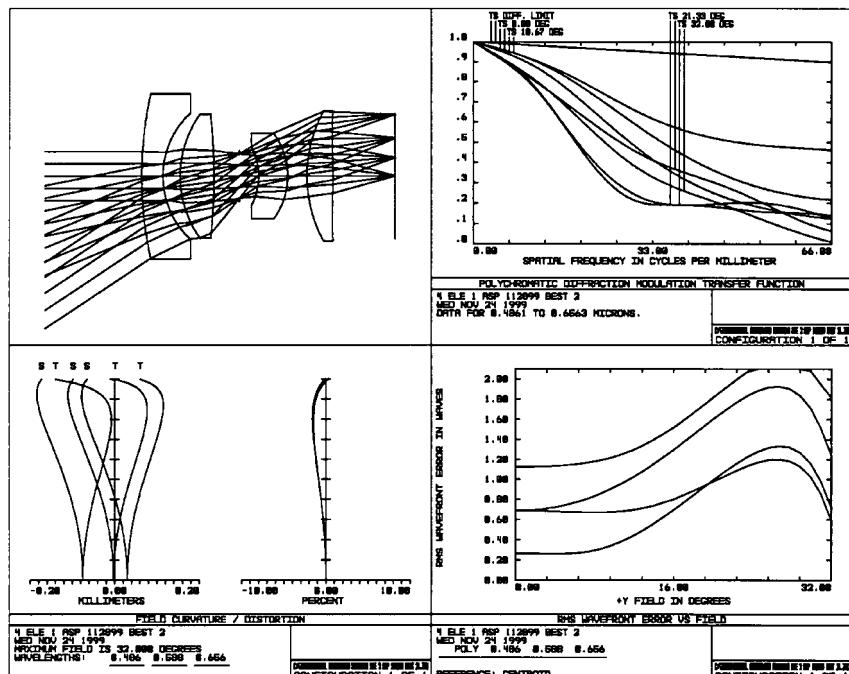
All lenses (Fig. 22.26) are made of high-index glass. Ray bending is smoother than in the previous configuration. Distortion is very low, but unfortunately, MTF is somewhat lower.

Design with Four Glass Elements, Three of Them Aspheric

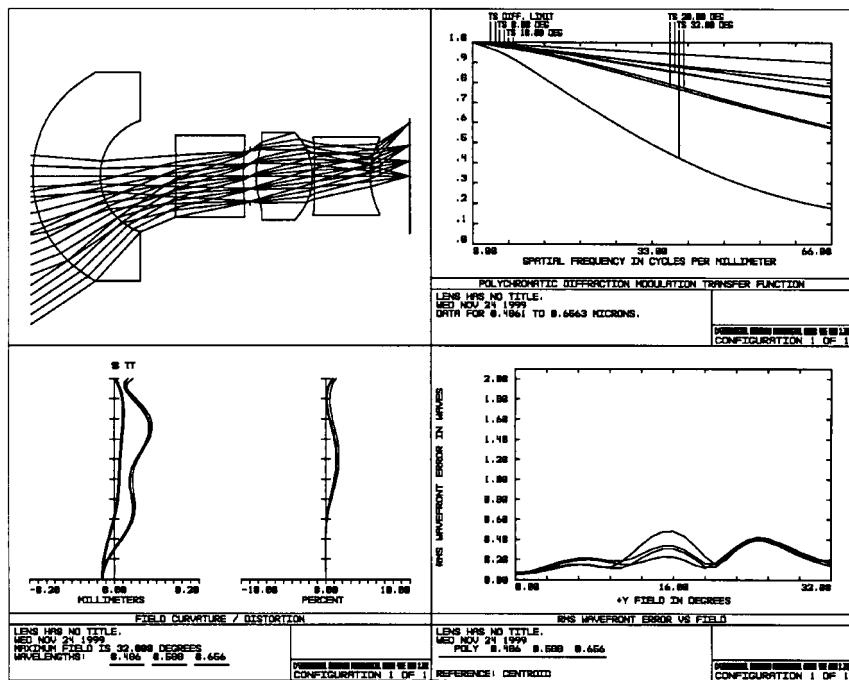
Figure 22.27 shows this configuration. This lens has very good performance. The first element may be difficult to manufacture cost-effectively, and note also that the angles of incidence on the sensor are high at the edge of the field.

Figure 22.26

Four Glass Lenses for
Digital Camera Lens,
Two of Them
Aspherics

**Figure 22.27**

Four Glass Lenses for
Digital Camera Lens,
Three of Them
Aspherics

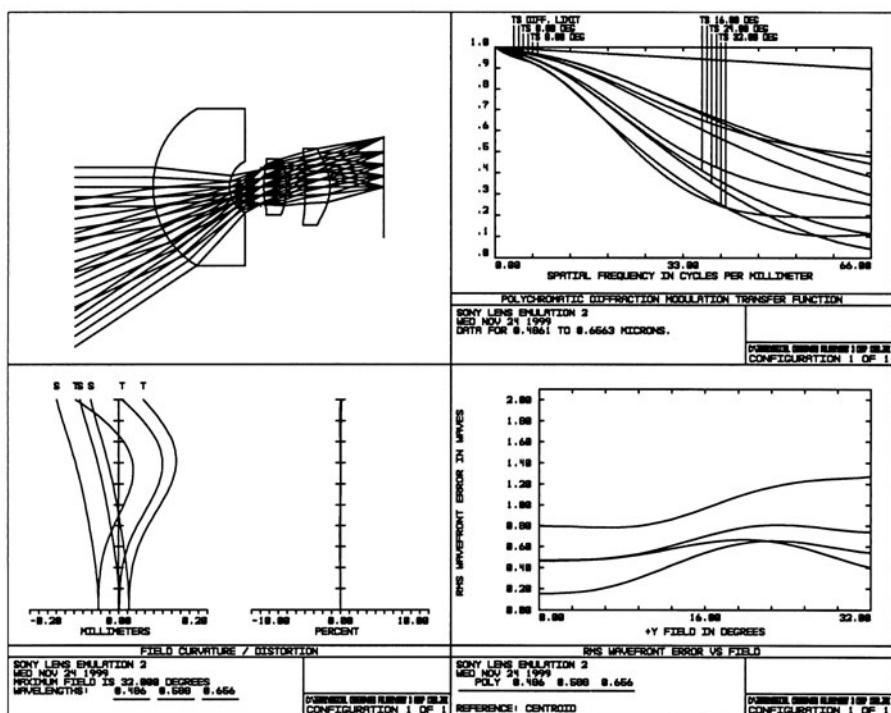


Design with Three Elements, Two of Them Glass Spherical and One Plastic Aspheric

The last configuration (Fig. 22.28) that we are going to show is a three-component lens. It has one plastic component, which is generally cheaper than a glass one. The performance is satisfactory, although the MTF is lower than in the previous four-component case. One parameter that should be controlled and the lenses compared to is the total track, which is the distance from the lens front surface to the CCD chip. We did not control this parameter, simply because we could not measure it in our camera. The last three-component lens shown has a shorter total track than the previous four-component lens, and it is preferable. It would be useful to investigate design forms with one diffractive surface, but we will stop at this point. The last two configurations shown could be good candidates for the final design. In the next step, a tolerance analysis, manufacturability, and cost analysis should be performed, and the final design chosen.

Figure 22.28

Two Glass and One Plastic Lens for Digital Camera



In our short exercise, we attempted to derive the design of a digital camera lens closest to the one in our camera. However, without destroying our camera, we may never know precisely what lens design form was used.

Binocular Design

This example is the design of a reasonable quality 7×50 binocular. The binocular should be low cost and as compact as possible. A binocular system is more compact with a Pechan than with Porro erecting prisms. However, a Pechan prism is more expensive, since it is a roof prism with a tight tolerance on a roof angle in the order of 2 to 3 arc-sec. That is the reason why we will design the system with a Porro erecting prism rather than a Pechan prism. The simplest objective is a cemented achromatic doublet. An $f/4$ achromatic doublet gives a reasonable quality image. This equates to a 200-mm focal length for the objective. The basic binocular specifications are listed in Table 22.5.

Good optical performance would require the system to have resolution in the exit pupil similar to the resolution of the human eye. If the eye can resolve 2 arc min/line pair, a good system should have, at the center of the

TABLE 22.5

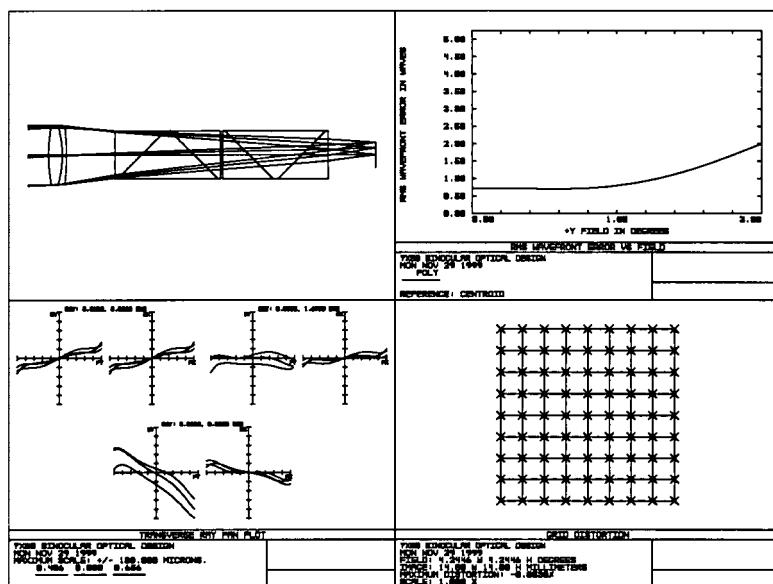
7×50 Binocular Design Example

Parameter	Specification
Entrance pupil diameter (mm)	50
Magnification	$7\times$
Objective focal length (mm)	200
Objective f /number	$f/4.0$
Eyepiece focal length (mm)	28.6
Full field of view (degrees)	6
Spectral range	Visual (C, d, F)
Distortion (%)	<12
Vignetting (%)	<30 at edge of field
Diameter of eyepiece assembly (mm)	<38
Eye relief (mm)	>23

field of view, not more than 2 min of spherical aberration and 2 min of chromatic aberration over the exit pupil size of 3 mm. For our system here, the optical performance can tolerate somewhat larger aberrations. The magnification of $7\times$ and the entrance pupil diameter of 50 mm give an exit pupil diameter of 7 mm. The design and analysis will be done for a 7-mm exit pupil diameter, and only in the final analysis, we will look at the system performance having an exit pupil diameter of 3 mm.

The design of the system starts with the design of the objective. First, we enter the doublet into the design program, and optimize it to minimize the aberrations and achieve the focal length 200 mm. In the next step we add two blocks of SK5 glass of the right thickness to simulate two right-angle prisms of the Porro system. In each right-angle prism there are two internal reflections, which can be simulated with tilted surfaces inside the glass block. The location of the prisms was chosen to leave minimum 40 mm from the exit surface of the second prism to the image plane. This is necessary to have enough room for the nose and comfortable resting of the binocular on the face. At this point, we decided to introduce a small amount of vignetting. The layout of the objective with the Porro prism and its performance is shown in Fig. 22.29. The second graph in Fig. 22.29 shows the rms wavefront error on a five-wave scale as a function of the field. The third graph shows the transverse

Figure 22.29
7 × 50 Binocular
Objective with Porro
Erecting System

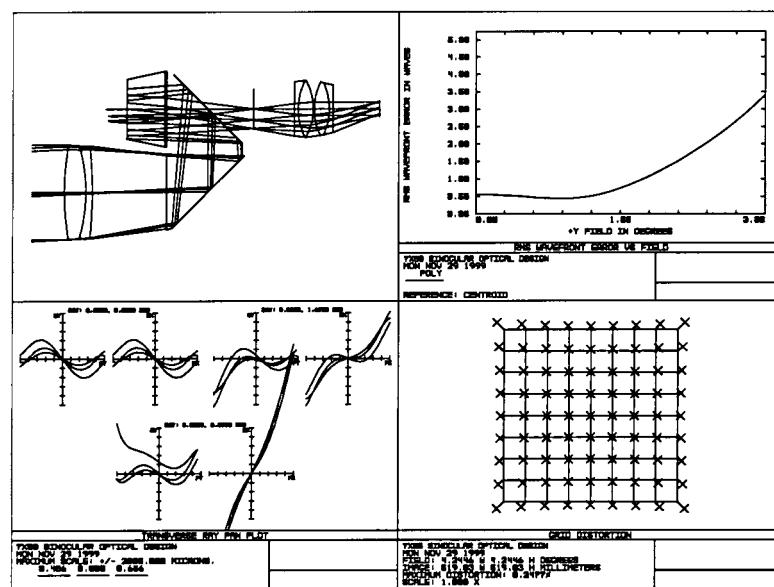


ray aberration curves on a $\pm 100\text{-}\mu\text{m}$ scale. There is some field curvature and astigmatism, as well as some lateral color aberration. The last graph shows a distortion grid, with practically no distortion.

In the following step we fold the prisms in four places where the reflections take place, check to see if the image plane is behind the plane in which the corner of the first right-angle prism is, and make adjustments to the location and size of the prisms, if needed. Now we can freeze the objective and add the eyepiece.

The full field of view of 6° in object space gives an apparent field of view to the user of $6 \times 7 = 42^\circ$, where 7 is the system magnification. A field of view of 42° is a little larger field than a symmetrical form of eyepiece is designed for, but it is a low-cost eyepiece which performs reasonably well, and we will design our system with it. We now add the symmetrical eyepiece and a paraxial lens in the exit pupil of the system. The optimized system with the symmetrical eyepiece is shown in Fig. 22.30. The eye relief is 25 mm. The second graph in Fig. 22.30 shows the rms wavefront error on a five-wave scale as a function of field of view. We notice some degradation in image quality toward the outer periphery of the field. The third graph shows the transverse ray aberration curves on the $\pm 2000\text{-}\mu\text{m}$ scale. These transverse

Figure 22.30
7 \times 50 Binocular
Design—Objective
with Folded Porro
Erecting System and
Symmetrical Eyepiece



ray aberrations are at the image of our paraxial lens, which was used to evaluate our afocal system. We used a paraxial lens of 1000-mm focal length. This means that the angular blur of 1 mrad coming into this paraxial lens corresponds to 1000- μ m blur in the image plane, or that our scale shows ± 2 mrad in the exit pupil. There is quite a lot of astigmatism at the edge of the field, otherwise performance is not bad. The last graph shows the distortion grid, with a maximum of 8.2% distortion.

We will now try a different form of eyepiece. If we start with a cemented doublet and two singlets, varying glasses and allowing the doublet to acquire a meniscus form, the resulting design has reduced astigmatism. This system is shown in Fig. 2231. However, this design has 12.7% maximum distortion, the diameter of the eyepiece assembly is larger, and the eyepiece is more expensive.

Let us go back and analyze the binocular design with the symmetrical eyepiece, with the 3-mm exit pupil diameter. The performance is shown in Fig. 22.32. We can see that the total blur on axis, including all colors, is smaller than 1 mrad. This is good, indeed. Some degradation in the image can be noticed in the last 25% of the field of view, which may be acceptable in our case.

Figure 22.31

7 × 50 Binocular Design—Objective with Folded Porro Erecting System and Four-Element Eyepiece

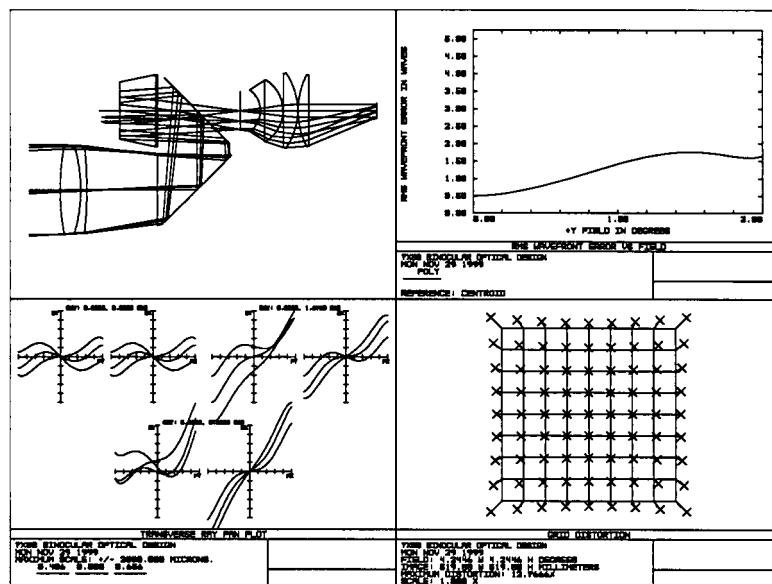
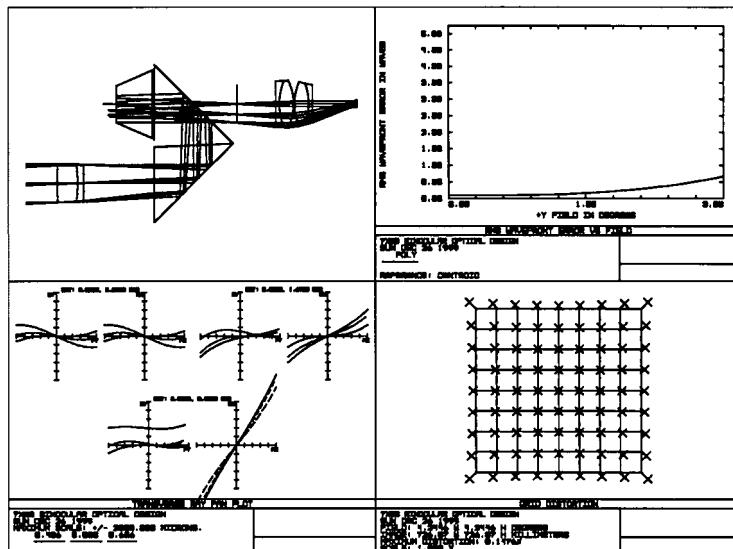


Figure 22.32
7 × 50 Binocular Design with Symmetrical Eyepiece Analyzed with 3-mm Exit Pupil



Parametric Design Study of Simple Lenses Using Advanced Manufacturing Methods

In order to illustrate the relative benefits of conventional as well as advanced manufacturing methods, it is often valuable to compare these methods parametrically. For lenses of the following specifications we optimized the performance of the design forms listed in Table 22.6. Also shown in Table 22.6 are the figure numbers:

- Clear aperture diameter of 12.5 mm
- Field of view of $\pm 2^\circ$
- Spectral band visible
- Materials BK7 and SF2 for doublets and BK7 for singlet, unless otherwise noted

The figures show for each design the following:

- Lens layout
- Transverse ray aberrations on a scale of $\pm 100 \mu\text{m}$
- Spot diagrams with a box width scale of $200 \times 200 \mu\text{m}$
- MTF plotted to a spatial frequency of 50 line pairs/mm

TABLE 22.6

Parametric Lens Design Examples

Lens Form	First f/Number	Figure Number	Second f/Number	Figure Number
Single element	f/2	22.33	f/4	22.34
Achromatic doublet	f/2	22.35	f/4	22.36
Achromatic doublet with aspheric surface	f/2	22.37	f/4	22.38
Single element with y^2 DOE	f/2	22.39	f/4	22.40
Single element with y^2 and y^4 DOE	f/2	22.41	f/4	22.42
Spherical acrylic Fresnel lens	f/2	22.43	f/4	22.44
Aspheric acrylic Fresnel lens	f/2	22.45	f/4	22.46
Gرادیم G1SFN gradient index glass	—	—	f/4	22.47
Planoconvex simple lens	—	—	f/4	22.48

Note that the scales for the optical performance are identical for all of the designs. While some of the data are off the scale, you will get a better understanding of the relative performance of each lens as it relates to the other approaches by using the same scale for the data.

Figure 22.33
f/2 BK7 Single Element

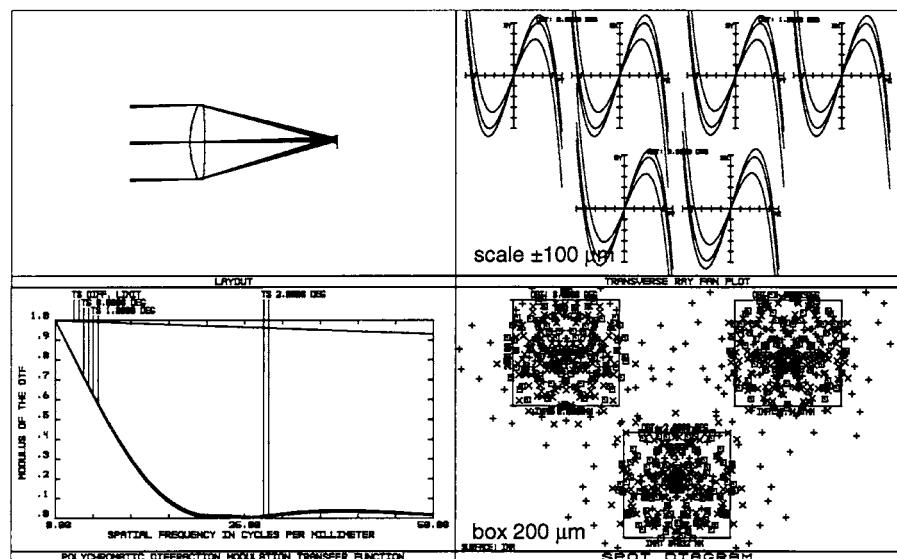
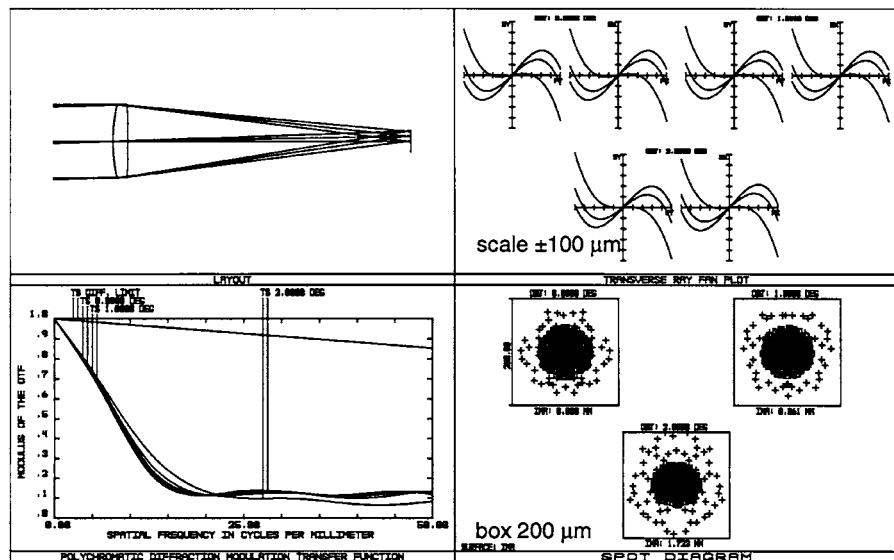


Figure 22.34
f/4 BK7 Single Element



The limiting aberrations in the *singlet designs* is spherical aberration. This is especially prevalent at the low $f/\#$ of $f/2$. At $f/4$ the spherical aberration is reduced to a level commensurate with the residual primary axial color. In both singlet designs the MTF is poor at 50 line pairs/mm, but this is not at all surprising.

Figure 22.35
f/2 Achromatic Doublet

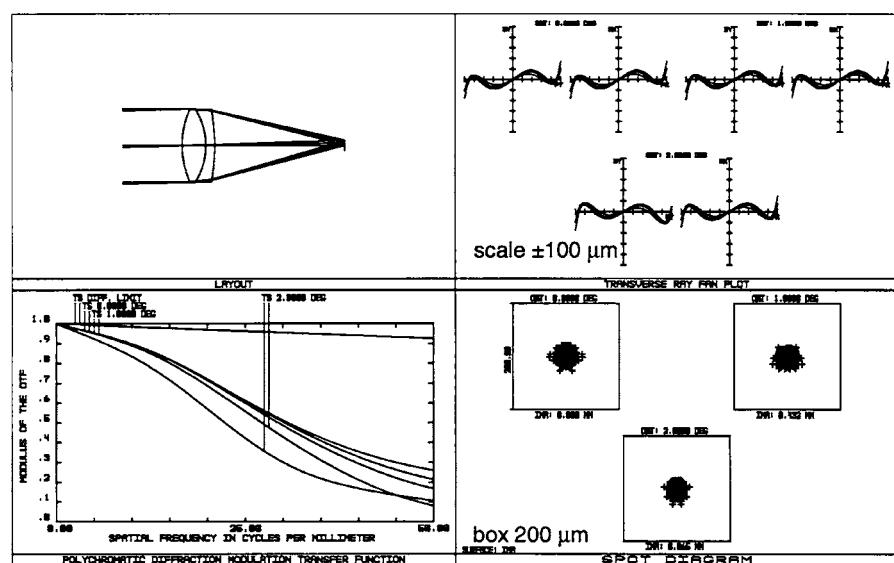
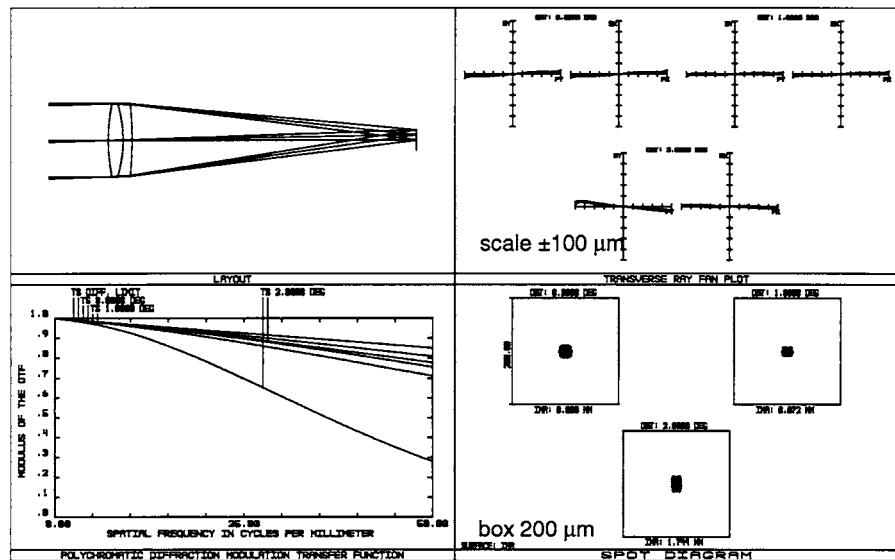


Figure 22.36
f/4 Achromatic Doublet



For the *achromatic doublet* the chromatic aberrations are reasonably well corrected with higher-order spherical aberration the limiting aberration for the $f/2$ lens. At $f/4$ the design is approaching being diffraction limited, except for the astigmatism residual at the edge of the field.

Figure 22.37
f/2 Achromatic Doublet with Aspheric Surface

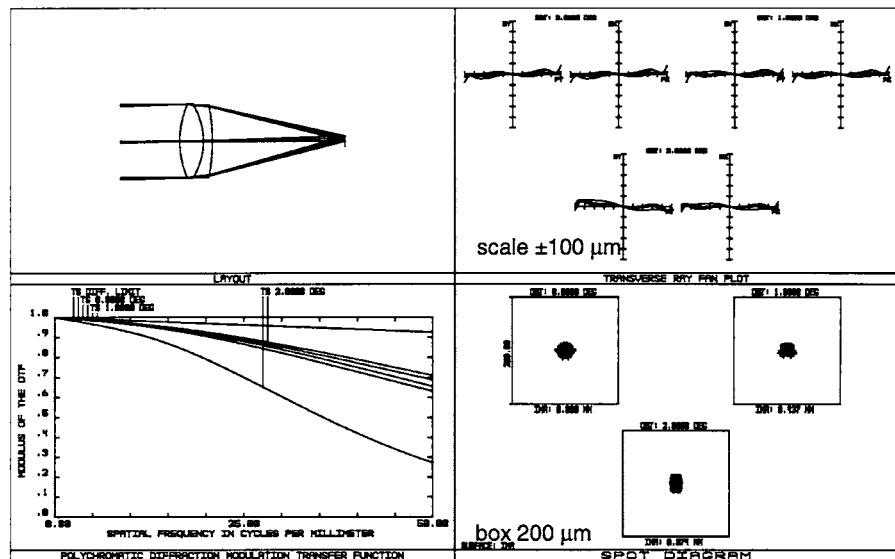
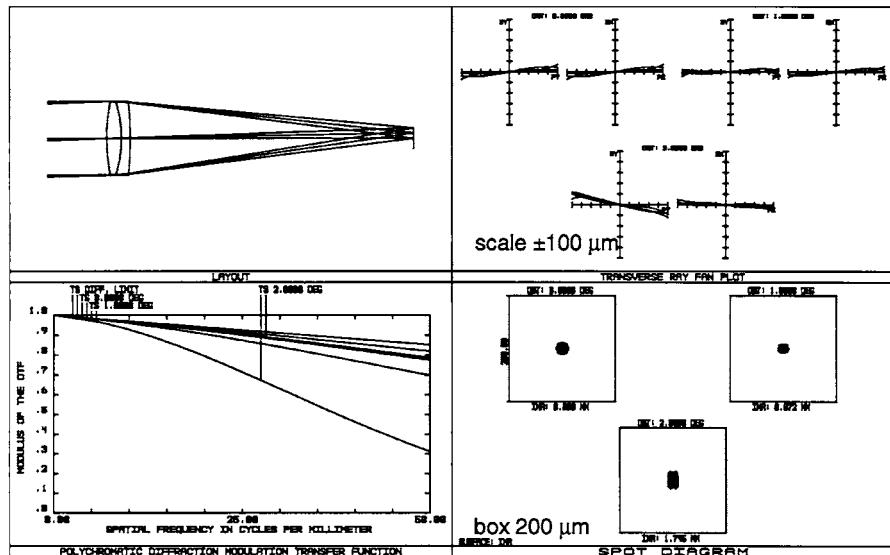


Figure 22.38
f/4 Achromatic Doublet with Aspheric Surface



An *aspheric surface on the achromatic doublet* allows for the correction of the residual spherical aberration with the astigmatism evident off axis. The chromatic aberrations are well corrected.

The *diffractive lens with a y^2 kinoform period surface* still has a residual of spherical aberration, and this is evident at both $f/2$ and $f/4$.

Figure 22.39
f/2 Single Element with y^2 Diffractive Kinoform Surface

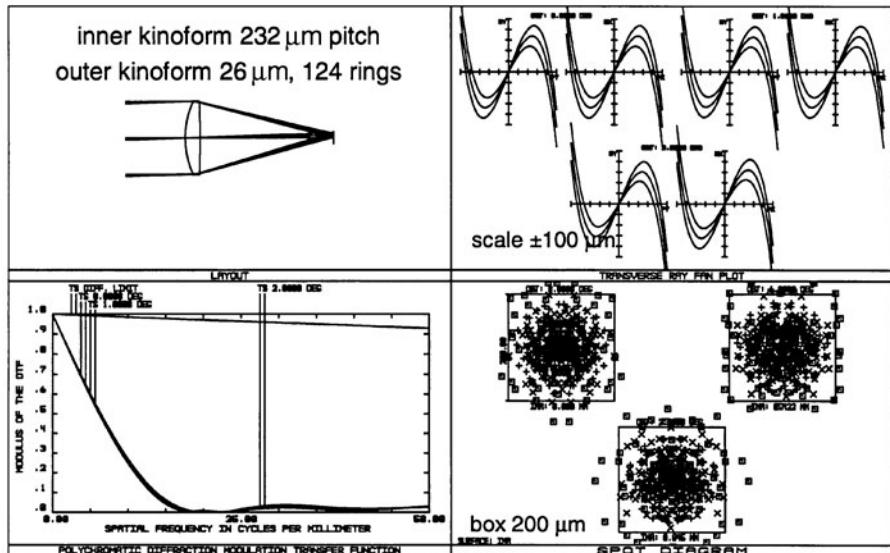
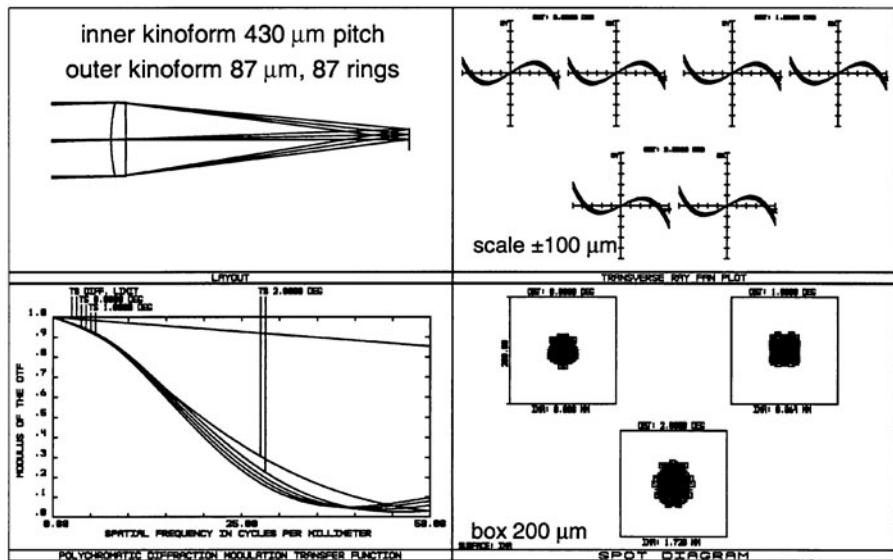


Figure 22.40
 $f/4$ Single Element
with y^2 Diffractive
Kinoform Surface



If we also allow, in addition to the y^2 , a y^4 *kinoform* period we can achieve a nearly complete control over the spherical aberration.

We have included *Fresnel lenses* for completeness. It is apparent that the residual spherical aberration of the spherical surface emulation of the Fresnel lens produces significant spherical aberration, in fact, far more than the equivalent conventional spherical single element.

Figure 22.41
 $f/2$ Single Element
with $y^2 + y^4$
Diffractive Kinoform
Surface

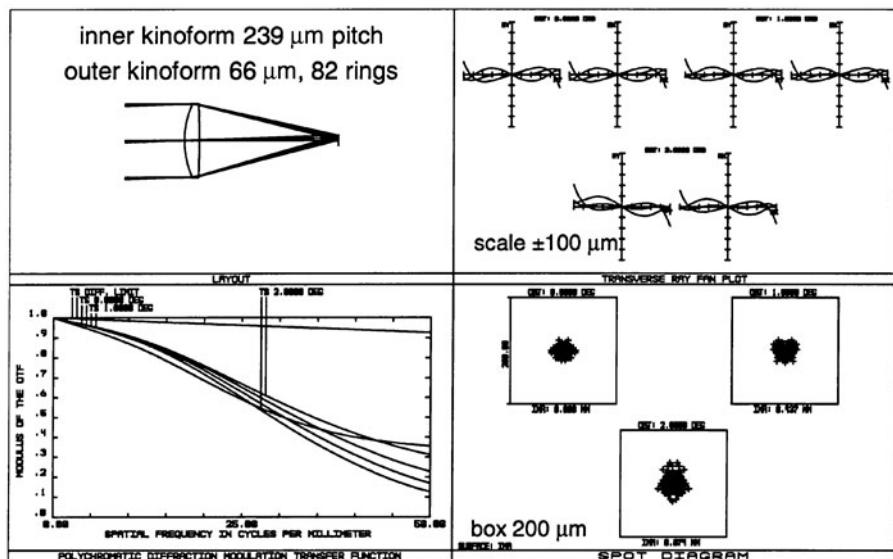
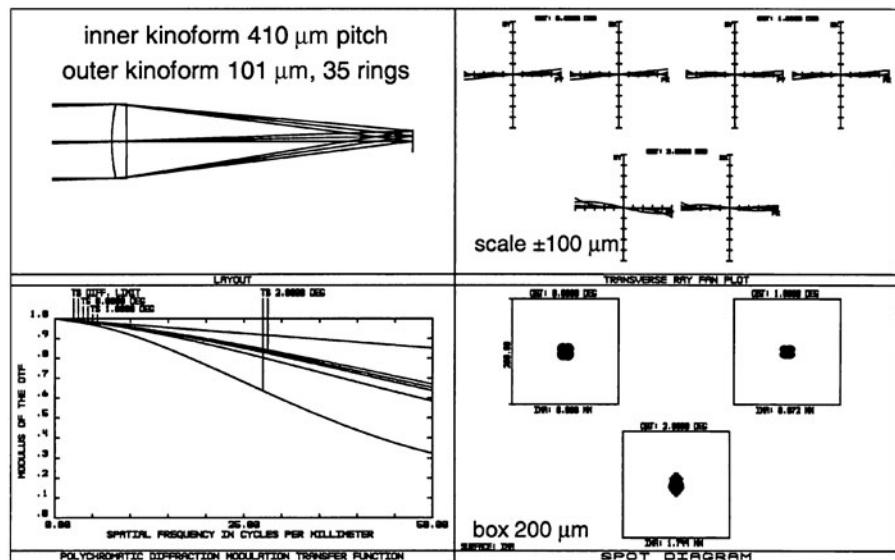


Figure 22.42
f/4 Single Element
 with $y^2 + y^4$
 Diffractive Kinoform
 Surface



The *aspheric Fresnel lens* is well corrected for spherical aberration; however, the primary residuals are primary axial color and coma off axis.

The *f*/2 planoconvex lens is shown for comparison with the *f*/4 Gradium lens.

Figure 22.43
f/2 Spherical Fresnel
 Lens

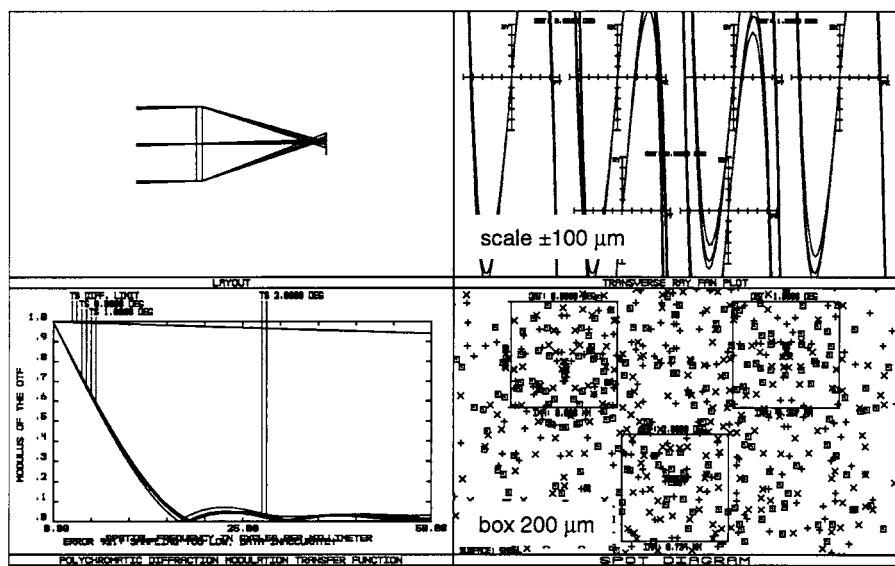
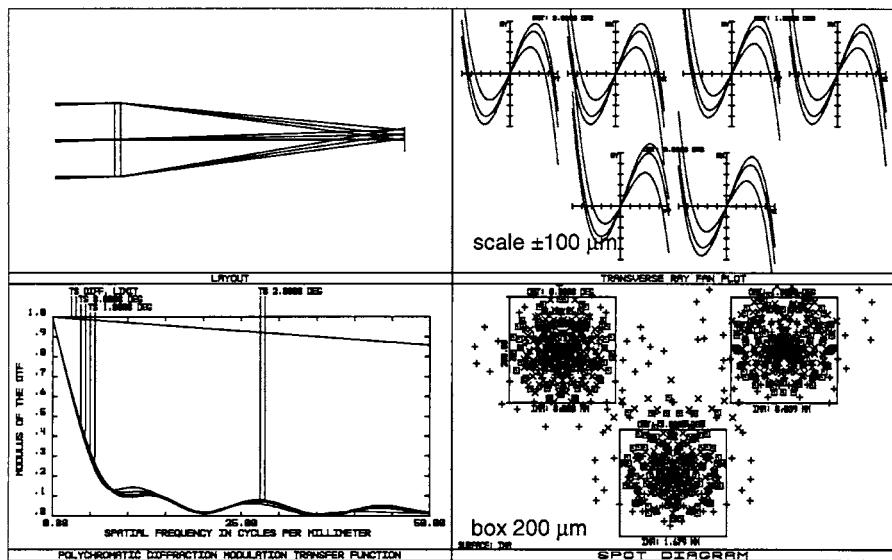


Figure 22.44
f/4 Spherical Fresnel
Lens



The *Gradium axial refractive index* gradient material is a material which permits most of the spherical aberration to be corrected while using spherical surfaces.

Figure 22.49 shows a summary of the rms blur diameters for each of the designs presented in this parametric study.

Figure 22.45
f/2 Aspheric Fresnel
Lens

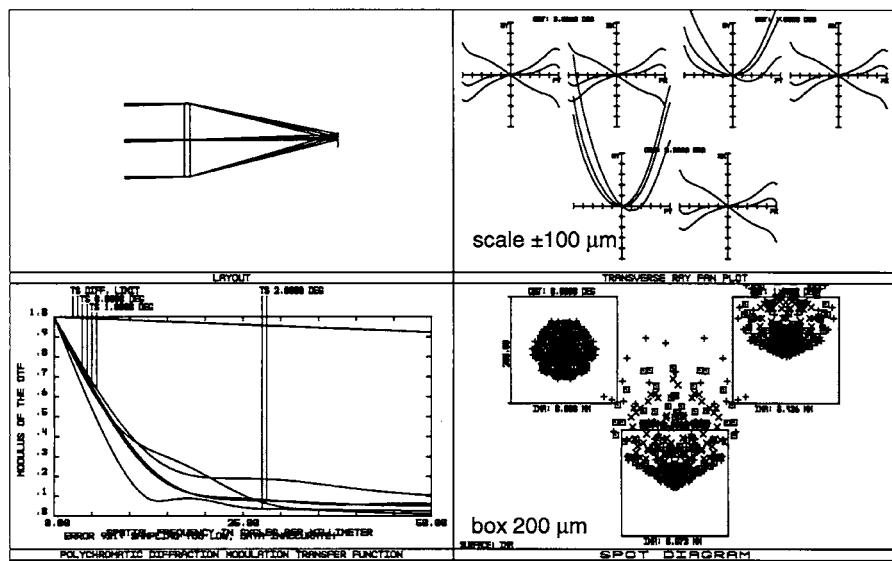


Figure 22.46
f/4 Aspheric Fresnel Lens

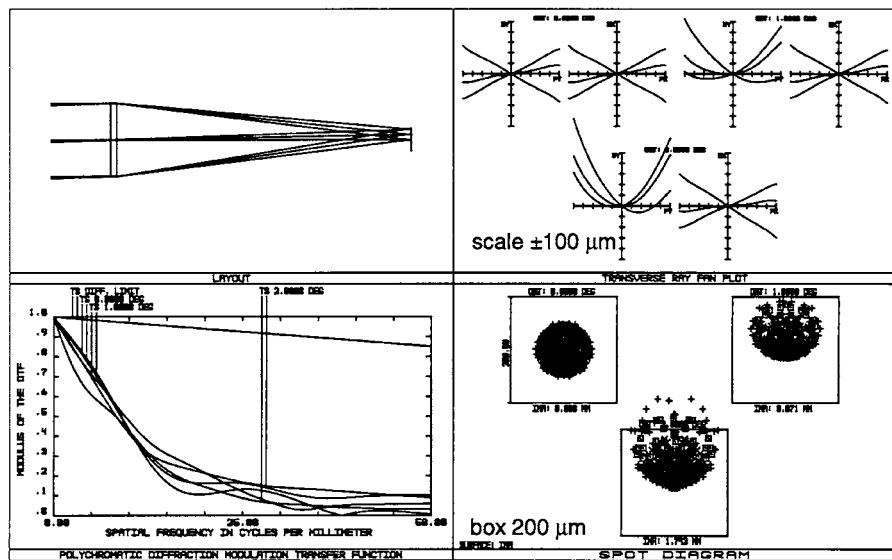


Figure 22.47
f/4 Plano BK7 Monochromatic

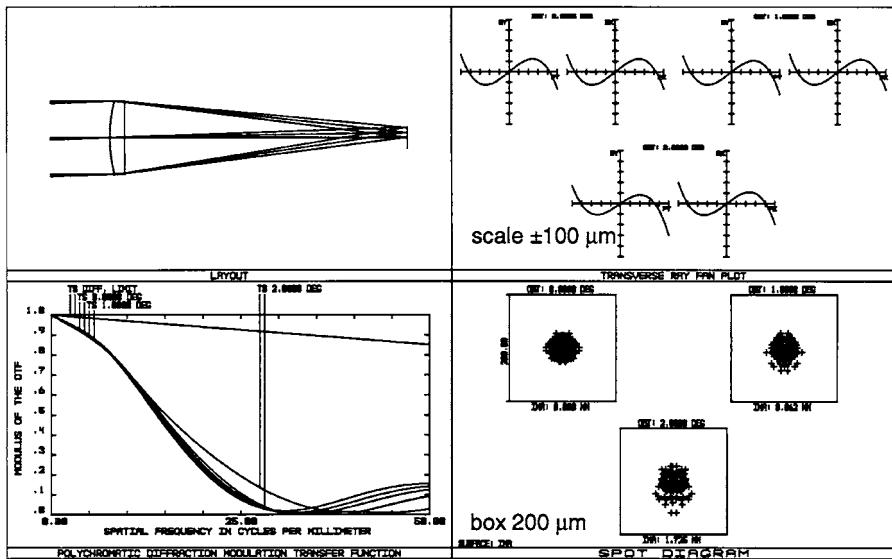
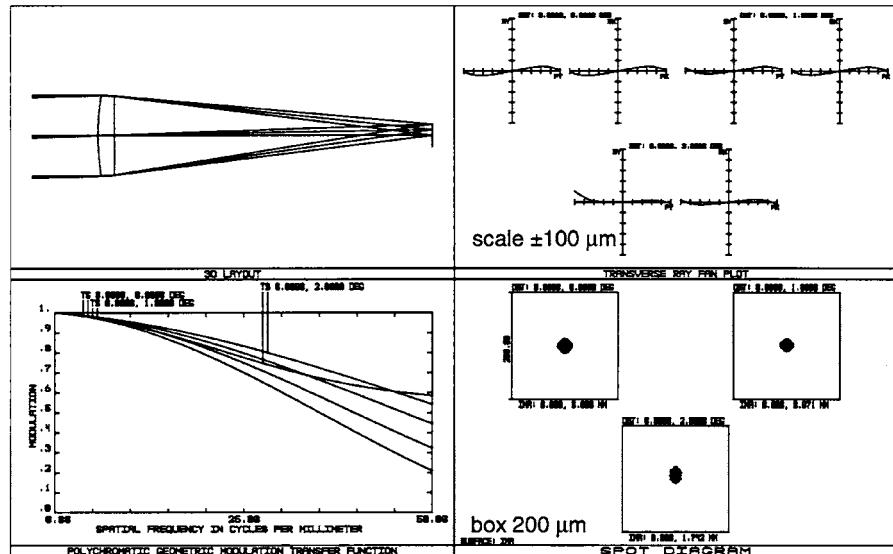


Figure 22.48
f/4 Plano
Axial-Gradient Index
Monochromatic



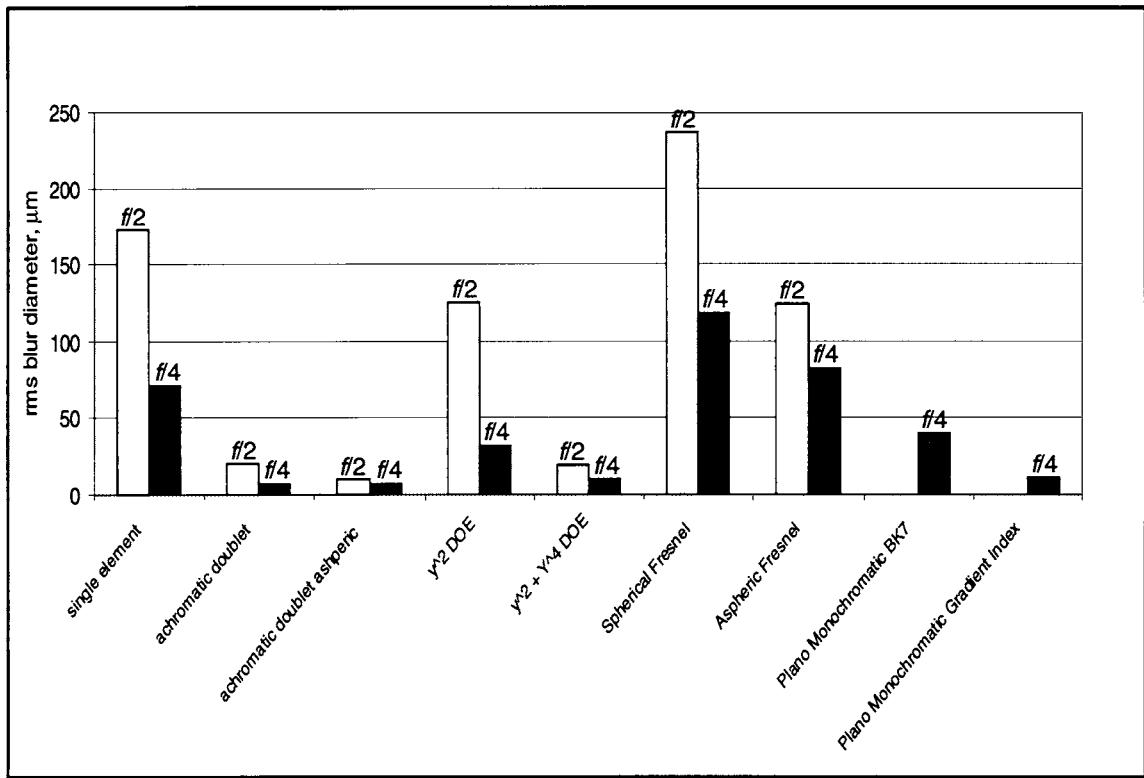
Design Data for Double Gauss

For reference we include here the optical design prescription data for the double gauss design example of the section “Double Gauss Lens Design” earlier in this chapter.

```

Title: Double Gauss Starting Design From Patent
System Aperture : Entrance Pupil Diameter = 25
Eff. Focal Len. : 50 (in air)
Image Space F/# : 2
Entr. Pup. Dia. : 25
Field Type: Angle in degrees
#      X-Value      Y-Value      Weight
1      0.000000    0.000000    1.000000
2      0.000000   11.000000    1.000000
3      0.000000   16.000000    1.000000
Vignetting Factors
#      VDX      VDY      VCX      VCY
1      0.000000  0.000000  0.000000  0.000000
2      0.000000  0.000000  0.000000  0.300000
3      0.000000  0.000000  0.000000  0.500000
Wavelengths : 3 Units: Microns
#      Value      Weight
1      0.486100    1.000000
2      0.587600    1.000000
3      0.656300    1.000000

```

**Figure 22.49**

RMS Blur Diameter for Different Design Approaches, f/2 and f/4 Lenses (Diffractive, Fresnel, and Gradient Index Lenses Are Single Elements)

SURFACE DATA SUMMARY:

Surf	Type	Radius	Thickness	Glass	Diameter
OBJ	STANDARD	Infinity	Infinity		0
1	STANDARD	Infinity	7.5		0
2	STANDARD	32.715	4.06	SSK51	25.34117
3	STANDARD	122.987	0.25		24.33643
4	STANDARD	20.218	7.3	SK10	23.12816
5	STANDARD	-112.78	2.03	F8	20.79916
6	STANDARD	12.548	5.08		16.57081
STO	STANDARD	Infinity	5.08		16.09063
8	STANDARD	-14.681	2.03	F15	15.37717
9	STANDARD	40.335	6.6	SSK2	16.93081
10	STANDARD	-19.406	0.25		17.96581
11	STANDARD	82.8556	4.11	SK10	19.73168
12	STANDARD	51.96156	32.24573		20.49336
IMA	STANDARD	Infinity			28.6088

Title: Double Gauss Final Design 12 Hour "Hammer" Optimization

System Aperture : Entrance Pupil Diameter = 25
 Eff. Focal Len. : 50 (in air)
 Image Space F/# : 2
 Entr. Pup. Dia. : 25
 Field Type: Angle in degrees

#	X-Value	Y-Value	Weight
1	0.000000	0.000000	1.000000
2	0.000000	4.000000	1.000000
3	0.000000	8.000000	1.000000
4	0.000000	12.000000	1.000000
5	0.000000	16.000000	1.000000

Vignetting Factors

#	VDX	VDY	VCX	VCY
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.006457	0.008781	0.090794
3	0.000000	0.008019	0.047112	0.204823
4	0.000000	0.000120	0.100113	0.337036
5	0.000000	-0.004495	0.248497	0.493301

Wavelengths : 3 Units: Microns

#	Value	Weight
1	0.486100	1.000000
2	0.587600	1.000000
3	0.656300	1.000000

SURFACE DATA SUMMARY:

Surf	Type	Radius	Thickness	Glass	Diameter
OBJ	STANDARD	Infinity	Infinity		0
1	STANDARD	Infinity	7.5		0
2	STANDARD	32.32399	5.345731	LAFN28	26.82534
3	STANDARD	84.92015	0.25		24.72447
4	STANDARD	19.13959	6.304279	LAFN10	22.61449
5	STANDARD	75.0351	2.03	LAF9	19.66355
6	STANDARD	12.66662	6.520931		16.11785
STO	STANDARD	Infinity	8.724626		15.09638
8	STANDARD	-15.0799	2.03	SF9	13.99697
9	STANDARD	-140.6069	5.060231	LAK33	17.28576
10	STANDARD	-21.26407	0.25		19.84303
11	STANDARD	91.13499	5.477296	LAK8	22.46901
12	STANDARD	-49.21186	25.4		23.51296
IMA	STANDARD	Infinity			28.43362

This page intentionally left blank

Optical Sensor Systems Modeling and Analysis

Introduction

Applications of optical systems vary hugely, a lowly spy-glass, a fully automated electro-optical sightline stabilized multichannel system, an optical storage read-write system, a directed illumination system, to name a few. In any optical system the optical components represent only a subset of the overall system. The overall system performance is therefore determined by more than just the performance of the optics. The optical components must be designed to perform in harmony with the rest of the optical system. Overdesign and over specification of the optical system should be avoided because this adds unnecessary cost and complexity to the optical design. Furthermore it is common that there are competing requirements within the optical system which need to be traded off. A common example is that volume and weight requirements will typically compete with lens resolution and light collection performance requirements.

In the case of a hand-held spy-glass the system performance depends not only on the quality of the lens, but also in clarity of vision of the person holding the spyglass and their ability to hold it steadily at the correct focal distance from the object.

A high-tech electro-optical system will also involve the interaction of a myriad of complex and interacting components all of which will

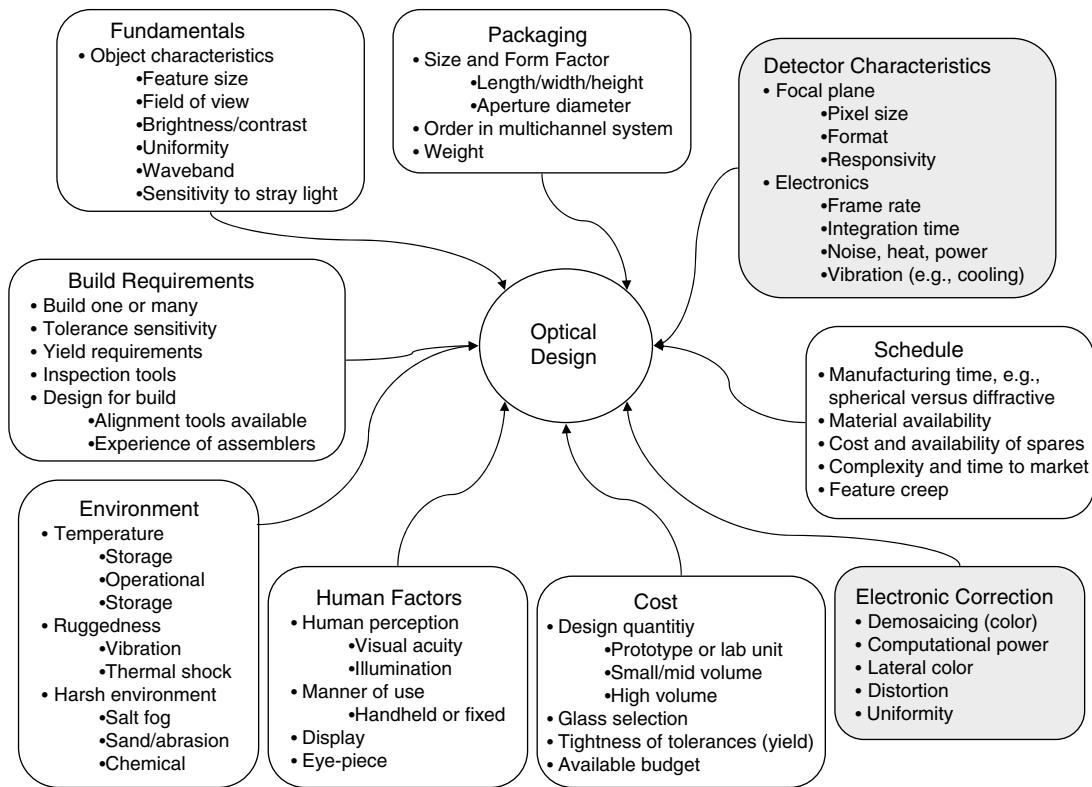
affect the final performance of the optical system. Examples may be the imaging performance of a lens train, the response of a camera (for example, charged-coupled device [CCD]), the read-out characteristics of the camera electronics, the mechanical stability of the platform on which the optics platform is supported, the signal strength of the object being viewed. Furthermore, the environment in which the system is used often needs to be considered as part of the optical system analysis. Effects such as temperature, humidity, atmospheric scattering and turbulence, and solar glare may need to be considered.

In any successful optical system, the influence of each aspect of the system on the final image must be understood in terms of what the system is being designed to observe or interact with. Of course, not all aspects of the system will affect the performance equally; some will be more important than others. Understanding and controlling the relative sensitivities of the different aspects of the system is important not only to achieve a technically sound optical system, but also to have a cost effective solution appropriate for the application at hand. To give an appreciation of the different factors which influence the optical design, refer to Fig. 23.1. This shows the various influences which can affect the optical system and hence the optical design path . . . and of course, not all of these are technical requirements. However, all of these factors can influence the technical direction taken in an optical system development.

This chapter discusses a broad range of topics which commonly affect performance and influence the design of a variety of optical systems. This chapter concentrates primarily, but not exclusively, on matters related to visible imaging optical systems. However, many of the topics discussed here need to be considered in other optical system types and wavebands. Particular consideration is given to the optical systems sensor, which is a strongly influences the overall optical systems design architecture. The highlighted groups on Fig. 23.1 indicate the areas most discussed in the following pages.

Image Formation

Generally, an image can be constructed from one of three techniques. These techniques are described following, and can be seen schematically in Fig. 23.2a, b, and c.

**Figure 23.1**

Considerations in Optical System Design

Single Point Detector Scanning

A single point detector can be scanned over the image plane (Fig. 23.2a), or more usually, the image plane can be scanned over a single point detector. By synchronizing the signal generated on the detector with the scanning optical system, an image can be created. Scanning in two axes is required to generate an image. This approach allows a simpler detector design, but at the expense of requiring a more complex opto-mechanical system. The single point scanning system creates an image, where each point across the image is observed at a slightly different time. This can cause image artifacts if the object is moving quickly compared to the scan rate. If the object is moving in the same direction as the scan direction

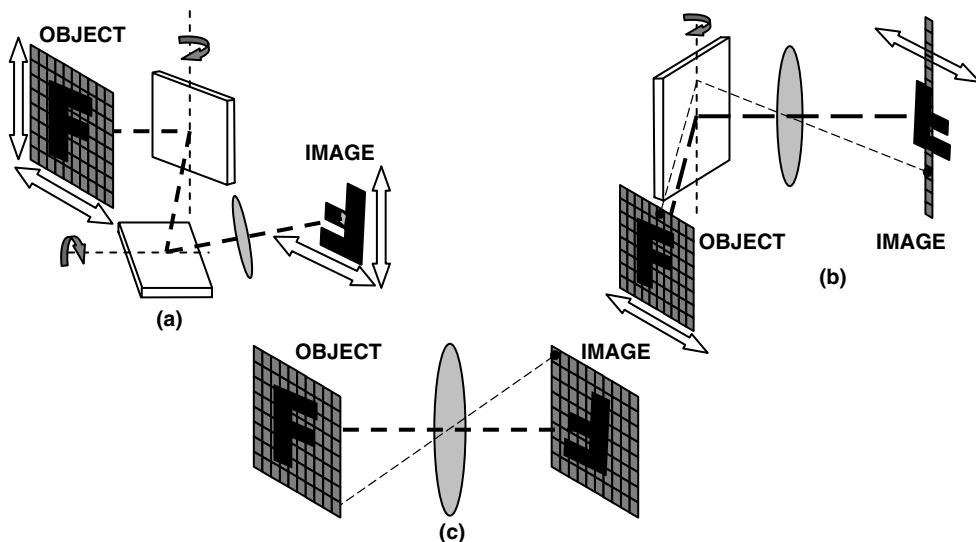


Figure 23.2

Image Formation with and without Scanning

then a smeared image will be obtained. Conversely, if the object is moving in the direction opposite to the scan axis, a compressed image will be obtained. In a single point scanning system, these image artifacts can occur in both the horizontal and vertical axes.

Linear Array Scanning

A one-dimensional array (or line) of detectors can be scanned over a scene to generate an image—see Fig. 23.2b. As shown in the figure, this can be achieved by use of a mirror. This approach is often used in airborne surveillance or satellite optical systems without the use of a mirror, where the scanning is achieved by the motion of the aircraft. Scanning in the direction of motion is known as a “push-broom” configuration. Scanning perpendicular to the direction of travel is known as “whisk-broom” scanning. Linear array imaging is commonly found in photocopiers and document scanners. The ability of the linear scanning system to detect a line of points simultaneously allows for longer integration times, which will improve the signal detected.

Two-Dimensional Array Imaging

A two-dimensional array of detectors requires no scanning (and therefore no moving parts) to form an image, hence the term “staring array” is often used (Fig. 23.2c). The use of an array of detectors simplifies the opto-mechanical design, because no scanning optics is required. An appropriately designed array imager may be operated in a snap-shot mode where all of the image points are recorded at the same instant in time. However, not all two-dimensional arrays operate in this way because of timing issues associated with reading the charge out of the detector elements; this is discussed further in the section below.

Detector Arrays

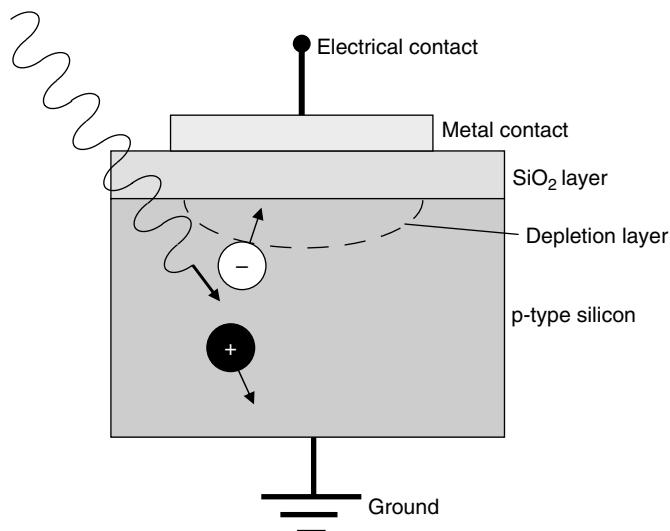
There are many issues surrounding the performance of an array detector which affect the final image quality. It is important that the optical engineer appreciate these sensor characteristics such that the optical design can be appropriately optimized. Array detector characteristics include sensitivity, noise, pixel size and resolution limitations, color attributes, frame rates, and connectivity.

Array Detector Descriptions

There are two commonly found types of visible array detector. These are CCDs and CMOS (complementary metal oxide semiconductor) arrays. Both technologies are essentially a closely packed array of semiconductor diodes. Light is detected when an incident photon excites an electron in the valence band of an atom up to the conduction band. This will only happen when the energy in the photon is greater than the energy band gap of the material. This is known as the *photoelectric effect*, the description of which earned Albert Einstein a Nobel prize in 1921. By collecting the freed electronic charge, the amount of light present can be measured.

CCDs are typically high performance, low noise detectors and are the preferred choice for high-end and scientific-grade applications. CMOS sensors offer a lower cost alternative to CCDs and whilst they have improved tremendously in recent years, they are unlikely to displace CCDs completely.

Figure 23.3
Simple Detector Structure



Differences in the charge collection architecture of CCDs and CMOS sensors give rise to different characteristics, however, the basic principle behind both detector types is the same. A photon generates an electron-hole pair. The charge is collected in a depletion region underneath the metal gate. This is shown in Fig. 23.3.

CCD Detectors

A CCD is an interconnected array of metal-oxide-semiconductor (MOS) capacitors, or pixels. Each pixel is a layer of SiO_2 is grown onto a p-type silicon substrate, onto which a metal gate is evaporated. The gate is positively biased with respect to the p-type silicon thereby creating a potential well between the SiO_2 layer and the p-type silicon; this can be thought of as an electronic charge trap. When a photon strikes the p-type silicon layer an electron-hole pair is formed. The electrons collect in the potential well and will remain there whilst the gate remains positively biased. The duration that the gate is positively biased is the charge integration time. The charge collected in the potential well is proportional to the flux falling on the detector. This charge must now be moved out of the potential well so it can be quantified.

CCD CHARGE READ-OUT It is worthwhile understanding the charge read-out mechanisms of CCDs. These may influence achievable

frame rate and integration time. This will of course affect the choice of fundamental optical parameters such as the *f*/number which can influence the whole optical system.

Charge read-out can be achieved by passing the charge from one pixel to the next, along to a common conductor pathway called the read-out register, a column found at the end of the row. The charge is transferred along the read-out register to a common node where the charge is converted to an electronic signal. During the read-out time, further charge cannot be collected. In the simplest form, charge integration can stop until all rows have been transferred to the read-out register. However, this would be slow, so there are two CCD charge read-out architectures commonly available, interline transfer and frame transfer.

Interline Transfer To speed up the read-out process, the charge can be passed to a buffer row or transport register. This is a row adjacent to a row of detector pixels which is shielded from the incident radiation by an opaque material (Fig. 23.4). Once the charge has been transferred to the transport register rows, it can be transferred to the read-out register column. Each transport register row is read to the read-out register column serially during the next integration period. This architecture is preferred in high-speed CCD cameras over the slower frame transfer architecture. A disadvantage of interline transfer is that resolution is reduced because every other row is opaque.

Figure 23.4
Electronic Read-out:
Interline Transfer

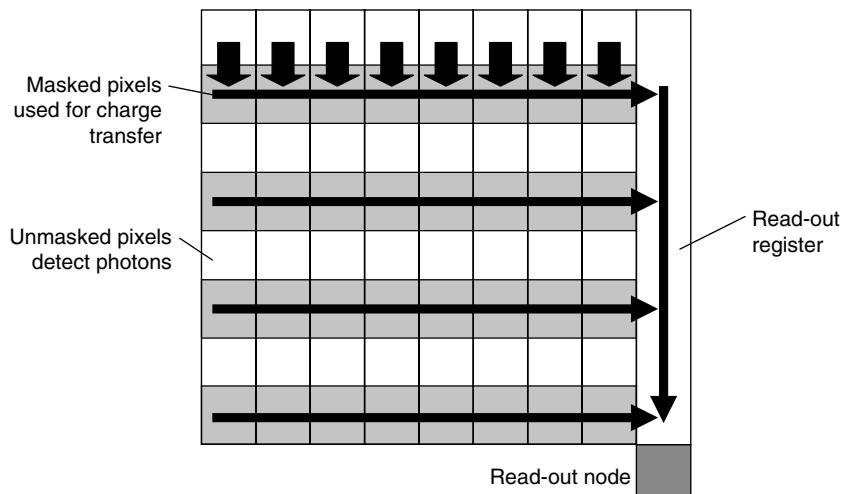
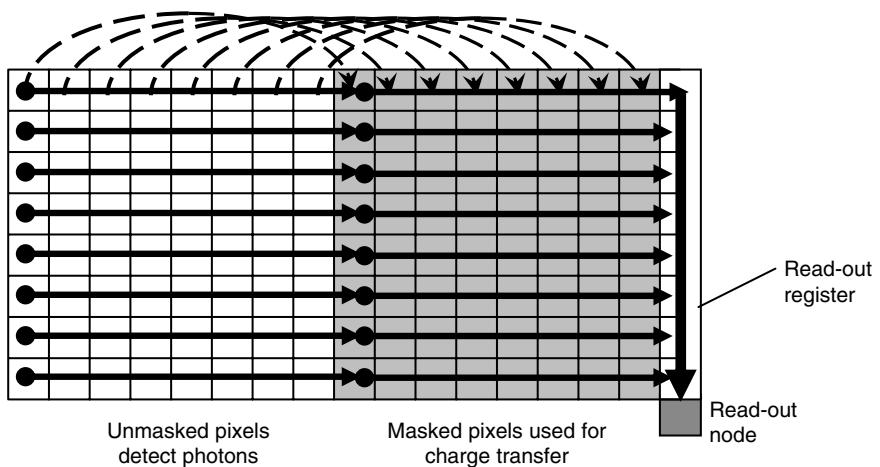


Figure 23.5
Electronic Read-out:
Frame Transfer



Frame Transfer To avoid the reduction in resolution found in interline transfer CCDs due to opaque rows, the row length can be doubled. The second half of each row is made opaque. The CCD therefore has two sections, the sensing section and the opaque storage section (Fig. 23.5). Following the integration period, the accumulated charge is transferred into the storage section on the same row. The storage section can then be read out into the read-out register whilst the sensing section is in the next integration period. This process is slower than interline transfer because it requires that charge be transferred along the full of the sensing section row before photon integration can start again. Frame transfer devices avoid the reduction in resolution, but the increased length of each row increases the cost of the detector.

Pixel sizes in CCD sensors is typically in the vicinity of $10\text{ }\mu\text{m}$; however, recent developments have reduced this significantly. CCD pixels as smaller than $2\text{ }\mu\text{m}$ have been developed. Scientific grade sensors used in applications where light to be collected from the object of interest is dim will use larger pixels, say $20\text{ }\mu\text{m}$. More area—more signal! The size of a pixel is the horizontal and/or the vertical dimension of a square or rectangular pixel. If you refer to a pixel's diagonal, make sure you indicate this. Generally the full width, full height, or full diagonal is referred to.

CMOS

CMOS sensors use the same principle for light detection as is found in CCDs, that is, photons separate electrons from atoms and the subsequent

collection of electrons allows the amount of light present to be measured. In the presence of an electric field, the freed electrons can be collected, and hence detection occurs. A CMOS sensor is an array of doped semiconductor diodes, usually of the p-n or p-i-n type.

Early MOS sensors were of the passive type. Each pixel was constructed from a photodiode and an access transistor. The accessed signal was passed down to an amplifier at the end of each column, a process which suffers badly from noise. CMOS sensors fabricated today are active pixel sensors (APS). Each pixel on an APS/CMOS sensor comprises a light sensitive region and active transistors. The active transistors buffer and amplify the accumulated photo-generated charge into a voltage, right at the pixel level. This significantly improves the noise performance. Passive pixel sensors are no longer used.

The APS architecture differs from CCDs where the charge must be read out of the pixel, into a read-out register and transferred to a common output node before being converted into an electronic signal. In an APS CMOS sensor, electronic signal can be addressed (read-out) pixel by pixel and in any order, unlike CCDs which can only be read out one row at a time. This offers a huge advantage for time-critical and region of interest (ROI) applications, where only certain pixels within the array need to be considered. Of course, the CMOS sensors need to have the appropriate supporting electronic and software interface to make use of this potential ROI functionality. The other advantage of converting the photo-charge to a voltage at the pixel level is that CMOS sensors do not suffer from *blooming*. This is caused when too much charge is collected on a pixel causing an overflow (saturation) of the pixel. The excess charge spills into adjacent pixels causing a signal smearing effect. Some CCDs do offer an antiblooming functionality by providing a drain to manage charge overspill. APS sensors on CMOS have been widely adopted because they can be made cheaply. CMOS sensors share the same processing steps required by computer logic and memory chip manufacture, for which there is a large high volume production infrastructure.

The power consumption of CMOS sensors is lower than in CCDs. This is largely because CCDs are typically driven with ± 5 to ± 15 V power supplies and have more off chip circuitry than CMOS sensors which are typically driven by 3 to 5 V supplies. CMOS sensors can be up to 100 times less power hungry than a CCD.

Disadvantages of the CMOS sensor are that on pixel transistors take up pixel real estate, which lowers the fill factor of a pixel. Instead of light hitting the photon receptive region of the pixel, some light will fall on the unresponsive transistor region. This fill factor reduces the

overall pixel sensitivity. Furthermore, because the charge to voltage conversion is on pixel, the noise filtering offered by the pixel is limited. On CCDs the charge to voltage conversion is bandwidth filtered to lower noise. On a CMOS sensors bandwidth filtering would require more on-pixel area which would reduce sensitivity, so this is not done. CMOS sensor on-pixel charge to voltage conversion operates with an open bandwidth which leaves CMOS sensors with more noise than found in CCDs. Whilst the noise characteristics of CMOS sensors have improved steadily they do not offer the same level of low noise of CCDs. CMOS sensor pixels are typically smaller than CCDs. CMOS sensors with a 4 μm pixel pitch are common and pixels smaller than 2 μm can be found.

Sensor Array Frame Integration Temporal Considerations

Typical read-out mechanisms for array sensors have been discussed above. A further issue worthy of consideration is to examine how the image is constructed in time. Again there are two primary cases to consider.

ROLLING INTEGRATION In a rolling integration architecture, a column is triggered to begin charge integration at a given point of time. All pixels in that column record (or integrate) over the same time period. At some point later, the next column begins integration. The time delay between one column starting integration and the next, may be approximately equal to the time required to read out a column in an interline transfer architecture. The advantage of a column staggered integration start time is that the frame rate of the camera can be increased. This is because the integration of the first column does not have to wait until all columns, from first to last, have been read out. The potential disadvantage of this is that each column records the object at a slightly different time period from the previous row. This may be especially significant in applications where the object is fast moving.

FULL FRAME (SNAP-SHOT) INTEGRATION In this case the integration time of every pixel is initiated at the same time. This means that the image is a snap-shot in time of the object. There is no difference in the time period recorded across the image generated. This is appropriate for the acquisition of images where the object is quickly moving. However, a snap shot frame will not be able to stagger the read-out from pixel to

pixel, or row to row, across the integration period as can be achieved in a rolling integration and read-out detector. This may reduce the maximum achievable frame capture rate. The read-out rate may, however, be less important in digital still photography, compared to creating a movie file.

Detector Response

QUANTUM EFFICIENCY (η) A electronic photo-detector converts light photons into electronic charge. Quantum efficiency (η) is a measure of the efficiency of this process and is expressed as the ratio of number electrons collected to the number of incident photons. Note that it does not include wavelength (or frequency) information.

RESPONSIVITY (R) The responsivity of detector includes photon energy. This is expressed as:

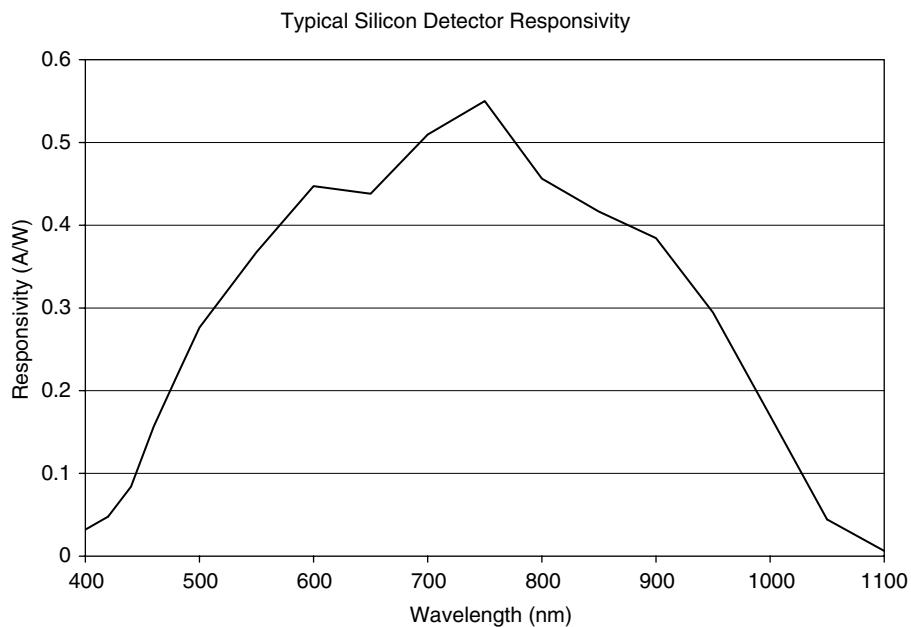
$$R = \frac{\eta e\lambda}{hc}$$

and has units of Coulombs/Joule. Over a given time interval we have units of Amps/Watt, where η is the quantum efficiency, λ is the wavelength of the photon, e is the charge on an electron, h is Planck's constant, and c is the speed of light. The responsivity versus wavelength curve of a typical silicon detector is provided on Fig. 23.6. Note that the sensor response to wavelength across the full visible spectrum (0.4 to 0.7 μm) and into the near infra-red (NIR), finally cutting off in the region of 1.05 to 1.1 μm .

Optical System Noise Characteristics

Noise sources often have to be considered because this will determine the signal to noise ratio the system achieves. If we understand the level of noise present in the system, we can then compute how much signal we require. The amount of light signal to be collected is one of the factors which drive the lens $f/$ # requirement; note that the other big factor in determining $f/$ # is the resolution requirement (see Chap. 1).

Figure 23.6
Typical Silicon Sensor
Response



There are many sources of noise within an imaging system, and the study of system noise can become very complex. A few of the more common noise sources are provided later on.

Noise sources can be combined as a root sum squared (RSS) because they are uncorrelated to each other. The total time averaged noise in an optical detection system can be expressed as in terms of a number of electrons, n :

$$\langle n_{\text{SYSTEM}} \rangle =$$

$$\sqrt{\langle n_{\text{SHOT}}^2 \rangle + \langle n_{\text{DARK}}^2 \rangle + \langle n_{\text{READ}}^2 \rangle + \langle n_{\text{RESET}}^2 \rangle + \langle n_{\text{QUANTIZATION}}^2 \rangle + \langle n_{\text{EXTERNAL}}^2 \rangle}$$

Care has to be taken in the combination of noise sources so that amplification effects are taken into account.

Dark Current Noise

Dark current is reverse bias leakage current which flows from the photodetector even when no photons are incident on the detector. Leakage current is inherent to semiconductor junctions. In a simple model, dark

current doubles for every 8°C temperature rise. The dark current gives rise to a dark current noise.

$$\overline{i_{\text{DARK}}^2} = 2eBI_{\text{DARK}}$$

where e is the charge on an electron and I_{DARK} is the dark current. This noise parameter is very important in scientific applications where long stare times are used. In scientific grade cameras, cooling is used to reduce the noise to very low levels.

Quantum Noise (Shot Noise)

The energy in a photon is greater than the energy in a thermionic emission, that is, $hc/\lambda > kT$, where h is Planck's constant, k is Boltzmann's constant, c is the speed of light, λ is the wavelength of light, and T is the temperature in Kelvin. The number of photons incident on the sensor will dictate the noise according to a Poisson distribution. Imagine if over a period of time a series of photons were incident on a detector one at a time. If the detected energy was plotted as a function of time it would spike every time the sensor experienced a photon *shot*. If we increase the number of photons, we will decrease the variance (that is, the noise) of the signal generated.

As a Poisson distribution we have:

$$\langle n_{\text{SHOT}} \rangle = \sqrt{\eta N_{\text{PHOT}}}$$

where η is the quantum efficiency and N_{PHOT} are the number of photons incident during the integration time.

Read-Out Noise (1/ f Noise)

1/ f noise is often referred to as flicker noise, where f is the frequency. The power in the noise signal is proportional to 1/ f .

Reset Noise Due to Thermal Noise (Johnson Noise)

Once the charge has been read from the output node, it must be reset to a known reference level before continuing.

The thermal vibration of ions within a conductor, such as a resistor, is proportional to temperature. The interaction of the vibrating ions and the free electrons gives rise to thermal noise which is described by:

$$\overline{I_{\text{THERMAL}}^2} = \frac{4kTB}{R}$$

where k is Boltzmann's constant, T is temperature (Kelvin), B is the electrical bandwidth of the system, and R is the resistance.

Quantization Noise

When the charge is read out it is converted from an accumulated charge, an analog signal, and into a digital signal. The analog to digital converter (ADC) will have a limit of how small an amount of charge can be discriminated. The magnitude of this gives the quantization noise. Quantization noise is dominated by the bit resolution of the ADC. The error is half of the least significant bit of the output. We can write that the electrical quantization noise for a rectangular pulse is:

$$\langle n_{\text{QUANTIZATION}} \rangle = \frac{LSB}{\sqrt{12}}$$

Note that the denominator does not vary with the number of bits.

Fixed Pattern Noise

Fixed pattern noise (FPN) is cause by background artefacts which appear on every image. This is not a temporally varying noise contribution, so is distinct from the other forms of noise discussed. FPN can be removed by background image subtraction, however, FPN may vary with environment, for example, thermal variations.

Bit Depth

Cameras are typically referred to as 8-, 10-, 12-, 14-, or 16-bit. This describes the number of levels into which signal counts can be recorded. The bit depth noise can be computed from $20 \log(n^2/2)$. The following table summarizes bit noise contributions:

Bits	Levels	SNR (dB)
8	256	42.1
10	1024	54.2
12	4096	66.2
14	16384	78.3
16	65536	90.3

However, care should be taken in simply opting for the highest bit camera. Pushing for a number of bits is not on its own helpful. If the noise of the camera is at the 10-bit level, the bottom 4 bits of the 14-bit camera will contain noise.

The bit depth typically relates to the well fill capability of the sensor. A larger pixel will likely have a larger well. Assuming comparable noise levels, this gives an opportunity to collect a larger signal and hence achieve an improved signal to noise ratio.

Improving on Pixel Fill Factor Limitations

To increase the fill factor, especially on CMOS sensors, microlens arrays are often used. Figure 23.7 illustrates the structure of a pixel with a microlens. This is further described by the image in Fig. 23.8, which shows a cross section of a real CMOS pixel structure.

The advantage of microlenses is that insensitive regions around the edge of the pixel can be avoided. In the case of a CMOS sensor, the sensing part of the pixel can be deep when compared to the surrounding layers in the structure. This creates a tunnel which the light must pass through. A microlens channels the light through that tunnel and onto the sensitive layer.

Microlenses also help to improve light collection if the pixel is less responsive near the edges or corner of the pixel. Similarly if there are gaps between the pixels, light can be redirected toward the center of the pixel to improve overall responsiveness to incident light.

The disadvantage of microlenses is that the optical designer must take into account the acceptance angle of the microlenses in the design. The angle at which the chief ray strikes the focal plane must be controlled to match the acceptance cone of the array. In Fig. 23.7a, where a microlens array is centered on every pixel, the optimal light collection is achieved

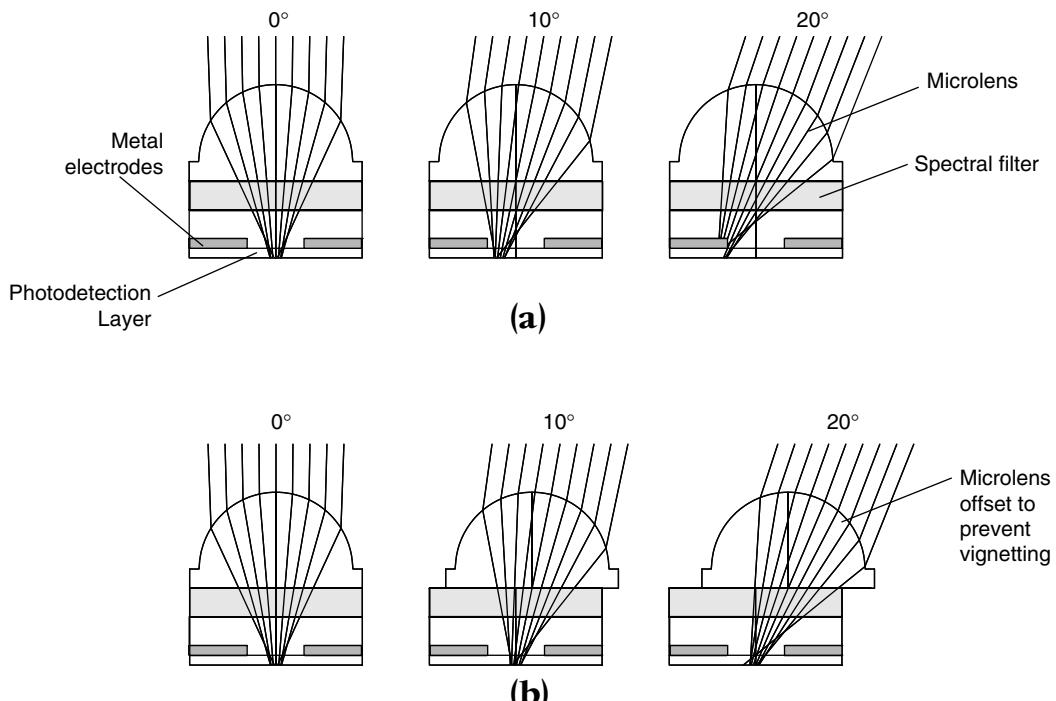


Figure 23.7 a and b

Microlenses Increase Fill Factor



Figure 23.8

Example of CMOS
Structure

with a telecentric design, that is, the chief ray strikes the image plane normally. The angle at which the chief ray hits the focal plane is often referred to as the “telecentricity angle.” This is something of a misnomer; a nonzero angle is really a measure of how nontelecentric the system is! Figure 23.7a shows that once the telecentricity angle exceeds 10° or so, vignetting of the incoming radiation on the metal electrodes buried in the pixel structure will occur; the signal detected will be significantly reduced. As the telecentricity angle increases to about 20° , nearly all the signal has gone.

Sometimes the microlenses’ pitch and the pixel pitch are slightly different. This means that the optimal acceptance angle is no longer telecentric. An example of this is provided in Fig. 23.7b. Once the telecentric requirements of the array sensor and the microlenses are understood the required angles can be optimized in the design. If matching is not achieved, then less light will be collected at the edges of the sensor than at the center. The result of this is a loss in brightness from the center of the image toward the corners. This is referred to as a reduction in relative illumination.

To avoid significant vignetting induced relative illumination roll off, the acceptance cone of the chief ray should typically be less than 10° or 12° from normal at the edge of the field, in most array sensors with microlenses. This margin will reduce with faster optical systems because a larger angular cone of rays will occur at the focal plane. This is one of those potential gotchas that needs to be checked for every new design!

Standard Sensor Sizes

CCD and CMOS sensor arrays are often described in terms of a dimension, usually in inches. This dimension can be quite misleading! This is because the description suffers from a legacy definition which describes the diameter of the sensor in terms of the diameter of a vidicon tube. Vidicon tubes are infrequently used in modern visible imaging systems. The diameter of the sensing area within the vidicon tube was approximately 66% diameter of the vidicon tube. This approximate factor still plagues the definitions today. Table 23.1 describes the relationship of descriptive dimension to the actual dimension.

The resolution formats typically used to describe sensors are provided in Table 23.2.

TABLE 23.1

Typical Detector Dimensional Formats

DIGITAL SENSOR FORMATS				
Type	Diagonal (mm)	Width (mm)	Height (mm)	Aspect Ratio
1/4"	4.5	3.6	2.7	4:3
1/3.6"	5.0	4.0	3.0	4:3
1/3.2"	5.68	4.54	3.42	4:3
1/3"	6.0	4.8	3.6	4:3
1/2.7"	6.72	5.37	4.04	4:3
1/2.5"	7.18	5.76	4.29	4:3
1/2"	8.0	6.4	4.8	4:3
1/1.8"	8.94	7.18	5.32	4:3
1/1.7"	9.5	7.6	5.7	4:3
2/3"	11.0	8.8	6.6	4:3
1"	16.0	12.8	9.6	4:3
4/3"	22.5	18.0	13.5	4:3

FILM FORMATS				
Type	Diagonal (mm)	Width (mm)	Height (mm)	Aspect Ratio
APD-H	34.51	30.2	16.7	16:9
APS-C	30.15	25.1	16.7	3:2
APD-P	31.66	30.2	9.5	3:1
35 mm	43.27	36.0	24.0	3:2

Pixel Pitch and Detector Nyquist Frequency

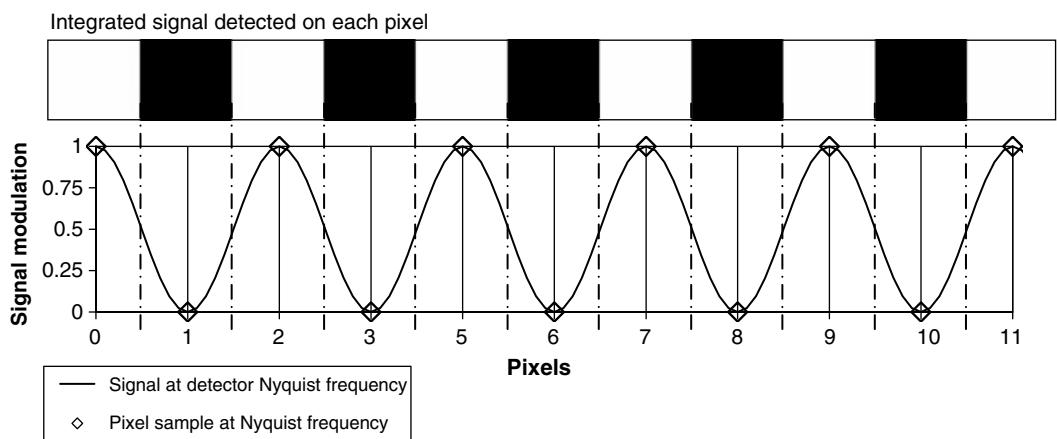
The detector Nyquist frequency is the maximum spatial frequency a pixelated detector will accurately detect without it being confused for another higher frequency. Imagine that a scene of black and white bars is being imaged onto the detector plane. The maximum observable frequency is where one pixel is illuminated with a white bar and the next is illuminated with a black bar, and so on (Fig. 23.9). If the bars are moved closer together, then the first pixel will begin to see a white bar and part of the

TABLE 23.2

Typical Standard
Sensor Resolution
Formats

Format	Pixels
QVGA	320×240
CGA	600×200
EGA	600×350
VGA	640×480
SVGA	800×600
XGA	1024×768
SXGA	1280×1024
UXGA	1600×1200
WUXGA	1920×1200
QXGA	2048×1536

next black bar; this causes the signal to become a little less white (slightly gray). The following will see slightly less of the black bar and a little more of the following white bar; it will be slightly lighter than black, that is, slightly gray. When this happens the frequency detected is not higher as it is in reality but becomes lower. At this point we are experiencing aliasing. This is discussed further in section Aliasing which follows.

**Figure 23.9**

Pixel Sampling of Sinusoidal Signal at Detector Nyquist Frequency

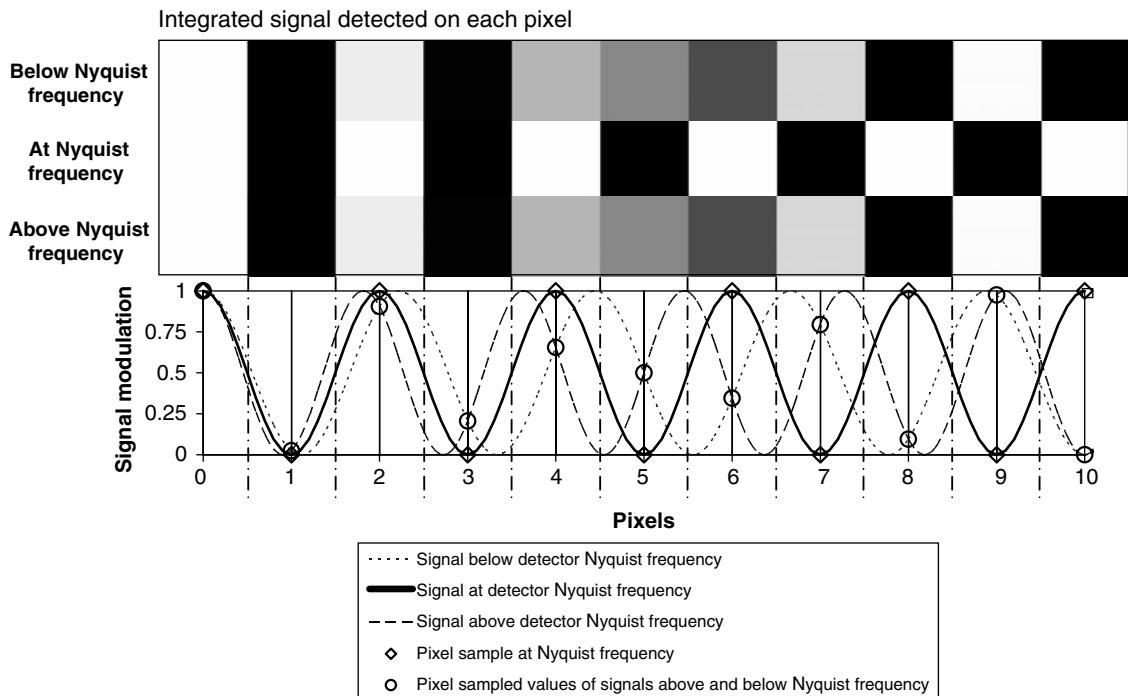
At the image plane, spatial frequency is defined in line pair per mm (lp/mm) or cycles per mm (cyc/mm). The Nyquist frequency is defined as $1/(2p)$, where p is the pixel pitch. Sometimes the spatial frequency is defined in object space; that is, the detector spatial frequencies are multiplied by the optical system magnification. As an example, a $10 \mu\text{m} = 0.010 \text{ mm}$ pixel pitch will provide a Nyquist frequency of $1/(2 \times 0.010) = 50 \text{ lp/mm}$ at the detector plane. Please be careful with this because we're talking about line pairs, that is, one white *and* one black bar. Spatial frequency may also be defined in terms of angle, usually in either object space. For example, a 100-mm focal length lens using a $10-\mu\text{m}$ pixel pitch sensor has a Nyquist frequency of 5 lp/mrad (line pairs per milliradian).

Aliasing

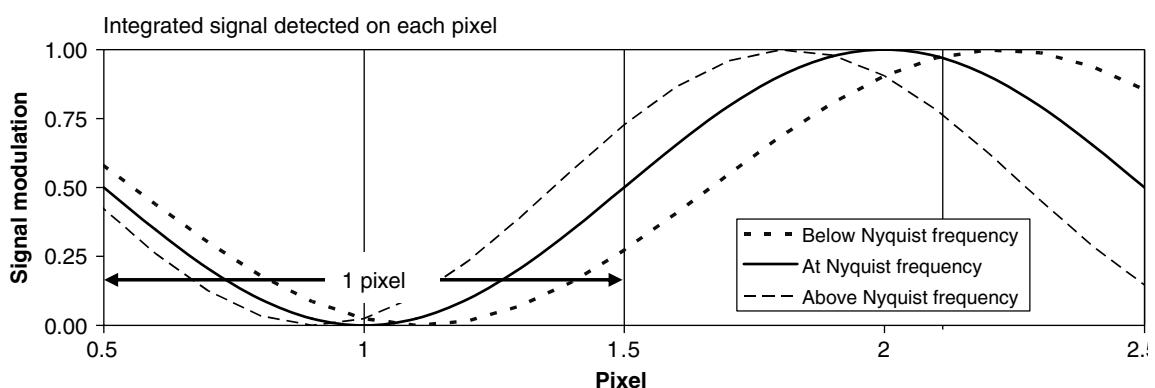
If a person has an assumed or alternate name, then that would be that person's *alias*. The same thing happens in the world of signals and frequencies! Aliasing happens when a pixilated sensor detects an incident spatial frequency to be a lower spatial frequency than it actually is. This occurs when the incident spatial frequency is above the Nyquist frequency of the detector. Spatial frequencies above the Nyquist frequency cannot be sufficiently sampled to correctly identify the true frequency. Consider Fig. 23.10, where a sinusoidal signal is sampled by pixels. The normalized signal on each pixel is shown as a series of white and black pixels representing the maximum and minimum signal which can be detected.

A continuous signal does not suffer from aliasing because it is not sampled. A pixilated detector, however, cannot be a continuous. By its very nature of operation, discrete samples are acquired across the image. This means that in any optical system using a pixilated sensor, aliasing will occur. There is no escaping it! What is important for the optical system is that the aliasing is understood and that it does not degrade the overall optical system performance. Of key interest to the system design is an assurance that the spatial frequencies of interest can be detected and that aliased spatial frequencies do not cause problems.

To help visualize aliasing, the case of three signals are considered. Firstly, one at the Nyquist frequency (f_{Nyquist}) and then two others, one slightly below ($f_{\text{Nyquist}} - df$) and the other above ($f_{\text{Nyquist}} + df$) the Nyquist frequency which is aliased. These sinusoidal signals are shown in Fig. 23.11. The sample points shown are at the center of each pixel. It can be seen that the signals above and below the Nyquist frequency ($f_{\text{Nyquist}} \pm df$) give the same values at the sample points. The normalized integrated signal

**Figure 23.10**

Pixel Sampling of Sinusoidal Signal near Detector Nyquist Frequency

**Figure 23.11**

Consideration of Signal Integrated on a Pixel

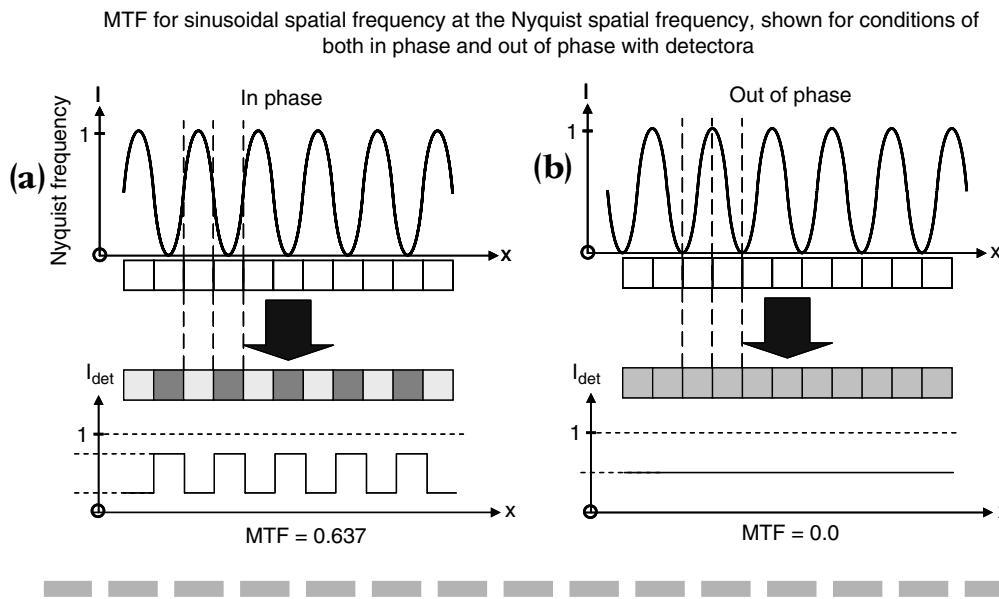


Figure 23.12 a and b

Signal In and Out of Phase with Detector at Nyquist Frequency

values for each pixel are shown, and it is clear that the signals above and below Nyquist give identical representations at the pixel level. Of course, only the lower spatial frequency is correctly detected. The higher frequency has been *aliased down* to a lower frequency.

It is worthy of note that the amplitudes the signals detected above and below Nyquist differ. This difference is small when the df is small. This can be seen in Fig. 23.12, where sinusoidally varying signals over 2 pixels are considered. Here the signal (area) under the lower frequency ($f_{\text{Nyquist}} - df$) curve integrates to lower and higher values on pixels 1 and 2 respectively, than does the aliased signal higher than Nyquist signal ($f_{\text{Nyquist}} + df$).

Detector Phase

The physical alignment of a spatial frequency distribution at the image plane with the detector pixels affects the signal which is generated on the detector. This will be considered by taking two cases. Firstly we consider a sinusoidal signal modulating at the Nyquist frequency of the detector, and secondly at the half Nyquist. Each case is considered in an in-phase and an out-of-phase condition.

MTF for sinusoidal spatial frequency at the half-Nyquist spatial frequency, shown for conditions of both in phase and out of phase with detector.

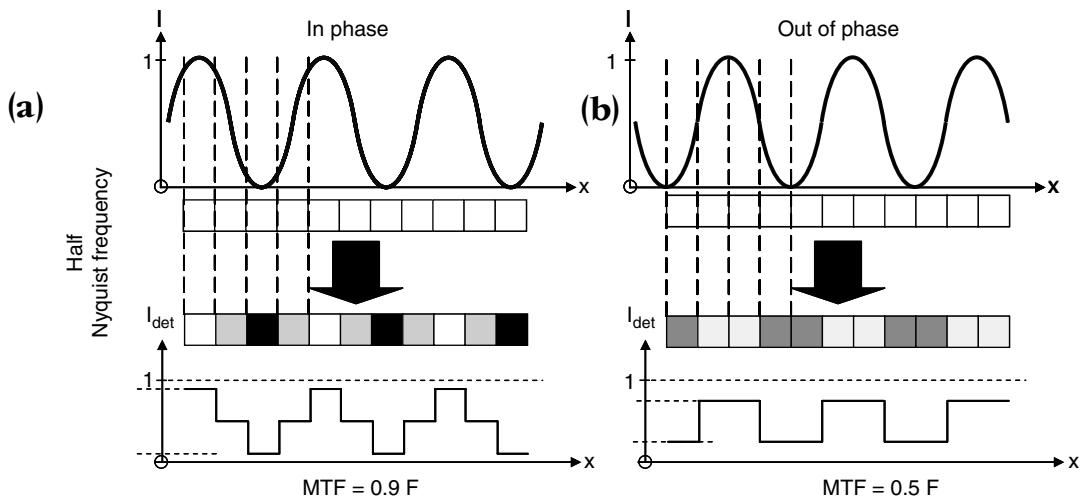


Figure 23.13 a and b

Signal In and Out of Phase with Sensor at Half-Nyquist Frequency

In Fig. 23.13a (*In Phase*), the peaks and troughs of the sinusoidal spatial frequency, the Nyquist frequency, are aligned with the centers of adjacent pixels. A series of black and white pixels are obtained indicating clear signal modulation. Now consider Fig. 23.13b (*Out of Phase*). Here a quarter wave ($\pi/4$ radian) shift of the spatial frequency (which is equivalent to a half pixel shift) has been applied. We have the same spatial frequency, same pixel sizes, but now we get no modulation at all! The bright and dark cycles of the modulation exactly match each pixel such that a mid gray level is obtained.

Now, consider a lower spatial frequency at the half-Nyquist. In Fig. 23.13a (*In Phase*) we find that good modulation is achieved. Again, with as little as half a pixel shift, we obtain a lower modulation in the out-of-phase condition, as shown on Fig. 23.13b (*Out of Phase*). Modulation at the half Nyquist is nonzero, but is significantly reduced from the in-phase condition.

Whilst it is unlikely that a real scene image will have most of the spatial frequencies out of phase with the detector, it is also unreasonable to expect that they will all be exactly in phase. Reality is somewhere in the

middle. The phase relationship of the signal and the detector is therefore clearly important if the total system performance is to be understood.

It should be noted that MTF values are provided in Figs. 23.12*a* and *b* and 23.13*a* and *b*. These values indicate the modulation due to the finite size of the detector pixel, and do not consider the sampling interval. For adjacent pixels, as in this case, the sampled signal is in fact the square of the value shown.

Pixel Sizes and Sampling Intervals

Both the size of the sampling area and the spacing between pixels influences the frequency response of the detector. In an ideal world a pixel would be an infinitely small point, with no spatial extent and pixels would be infinitely close together, thereby allowing a continuous function to be obtained. However, in real life pixels need to have a finite size and a separation distance and discrete sampling is required.

Both the finite extent of the pixel and the sampling distance between centers can be thought of as a finite-sized sampling window. As such, both the pixel and the sampling interval have a frequency response of their own. In both cases the frequency responses can be obtained by taking the Fourier Transform of a rectangle, which is a sinc function [$\text{sinc}(\pi x) = \sin(\pi x)/(\pi x)$]. In the case of the pixel, $1/f$ is the size of the pixel and in the case of the sampling interval, $1/f$ is the separation between sampling centers. The two MTFs are combined by multiplying them together to obtain the combined MTF response of the detector, including finite pixel extent and sampling.

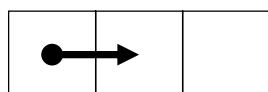
To get an appreciation of the frequency response of a pixel, and the sampling interval, three cases are considered. These are shown in Fig. 23.14. Firstly in Fig. 23.14*a*, the center to center pixel spacing equals the pixel width. This is common to monochrome array sensors. In Fig. 23.14*b*, the case where the active area of the pixel is smaller than the pixel to pixel separation is considered. Here the pixel to pixel separation remains the same as in case (a), but the pixel width is cut in half. In Fig. 23.14*c*, a full-sized pixel as in Fig. 23.14*a* is considered, but with pixel to pixel centers twice the pixel width. This would be typical of the raw frequency response of the blue or red pixels on a color array with a Bayer pattern (this is discussed further in section on Bayer patterns).

In Fig. 23.15, MTF and pixel sampling, the frequency responses to cases 23.14*a*, *b*, and *c* are provided. Firstly, the frequency response of a single

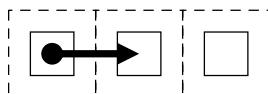
**Figure 23.14**

Pixel Size and Sampling Intervals

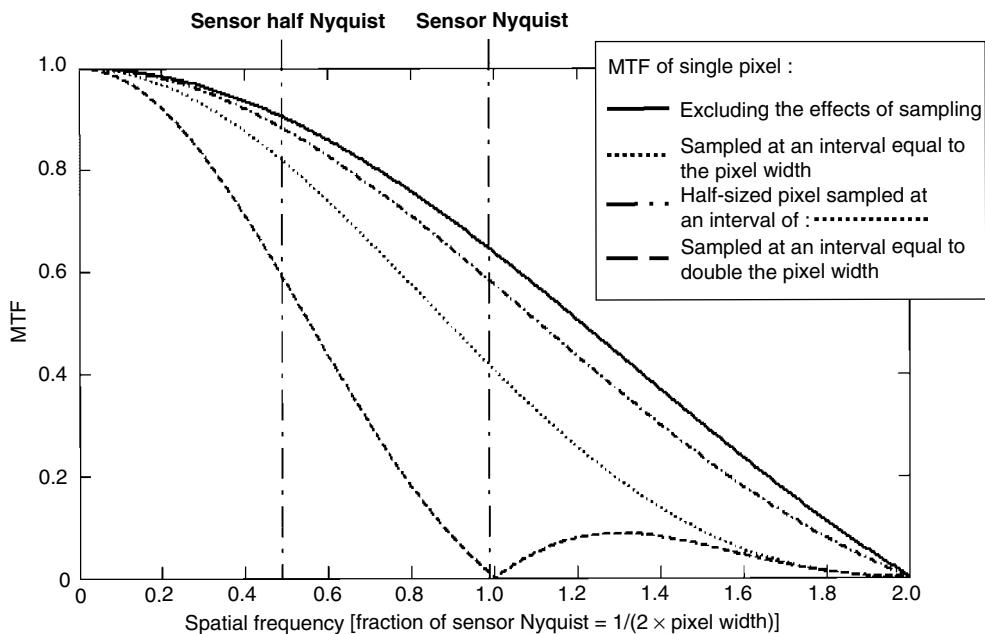
(a) Pixel sampling equals pixel width



(b) Pixel sampling as (a), and pixel width half of (a)



(c) Pixel sampling 2x (a), and pixel width equals (a)

**Figure 23.15**

MTF of Pixel and Pixel Sampling



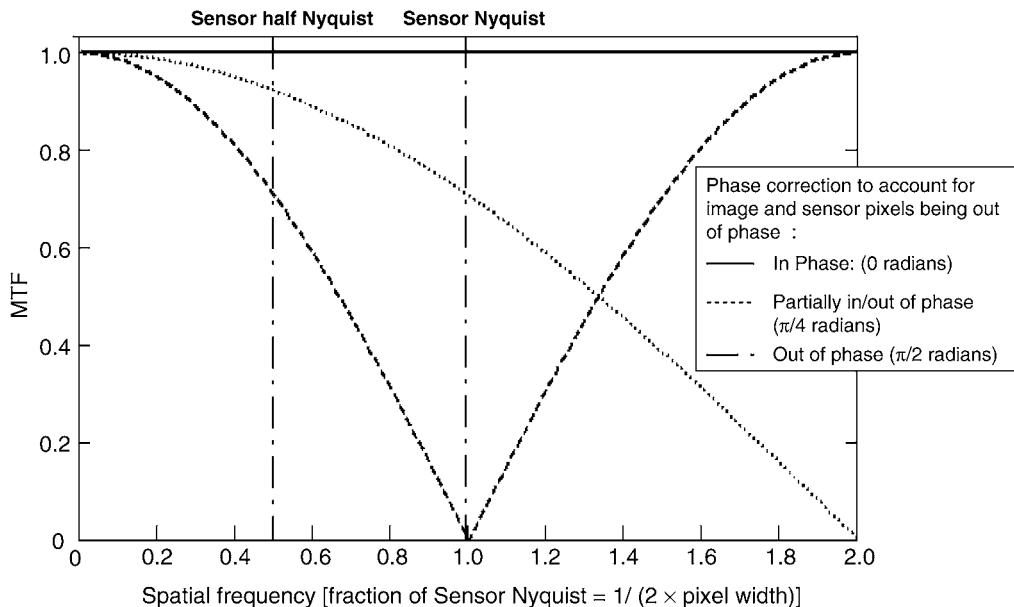
pixel is shown without sampling considered. The next case (Fig. 23.14a) is effectively a $\text{sinc}(\pi f)^2$ function. The reduced pixel size of Fig. 23.14b, including sampling, gives a better response. This is in effect $\text{sinc}(\pi f) \cdot \text{sinc}(\pi f/2)$; the narrower spatial extent of the pixel gives a broader frequency response, that is, a higher MTF. In the final example, Fig. 23.14c, of the full sized pixels with twice the gap between the pixel, the frequency response falls off further. This is reasonable, as it essentially represents poorer spatial sampling than the other two example cases. This is effectively a $\text{sinc}(\pi f) \cdot \text{sinc}(\pi 2f)$ response where we now find that the frequency response reaches zero at the detector Nyquist frequency. It is noted that the frequency response bounces, reaching zero at the Nyquist frequency. This bouncing of the MTF curve indicates that higher spatial frequencies will be detected, but will be aliased. Furthermore, the bouncing indicates that there will be a π phase reversal for aliased frequencies. Now, not only will frequencies be aliased, but the whites will become blacks and vice versa. The amplitude of the modulation is, however, reasonably low (max is about 10%), so aliasing should have poor contrast in this case. In these examples, all signals are in phase with the detector.

Pixel Sizes, Sampling Intervals, and Phase

As shown earlier, the impact of phase between the image and the pixels can be significant, so this is now considered in the context of finite pixel detection sampling and sampling intervals. Phase can be represented by an additional cosine function. When in phase, a 0 radian phase shift is applied. A $\pi/2$ radians phase shift represents an out of phase condition, and $\pi/4$ radians is a representative compromise condition. The impact of just that phase is shown on Fig. 23.16. These three-phase cases are now applied to the earlier example Fig. 23.14a of a single pixel with centers separated by the pixel width.

Figure 23.17 (MTF of Pixel and Pixel Sampling Considering Phase) will degrade considerably as it is moved to an out of phase condition. It is of interest to note that the partial in/out phase condition drops the MTF by about 10% (absolute) which is about 20% of the existing modulation. The remaining 80% of the available modulation falls away with further phase alignment errors.

It should be noted that in an optical system that is not detector limited, we will not suffer from aliasing because of the detector sampling. If the lens on the optical system is limiting the performance, then it will

**Figure 23.16**

Cosine Phase Factor to Account for Phase Between Image and Detector Pixels

effectively be acting as a low pass spatial filter; high-spatial frequencies are simply not imaged on to the focal plane. This phenomenon is sometime deliberately used to prevent aliasing. Optical frequency filtering is sometimes employed to prevent detector aliasing. A commonly employed example is the use of the optical low pass filter (OLPF).

OLPF (Antialiasing Filter)

An OLPE also know as an antialiasing filter, the OLPF is an optical low pass spatial frequency filter which serves to remove spatial frequencies which would otherwise alias on the detector. This is achieved by splitting a ray just in front of array detector into several spots. Typically one spot is split into four spots. This is in effect a deliberate and controlled blurring of the spot at the focal plane. Yes—after all the hard work of designing a top performing lens, we're throwing away resolution by deliberately blurring it! The separation of the beams controls the spatial

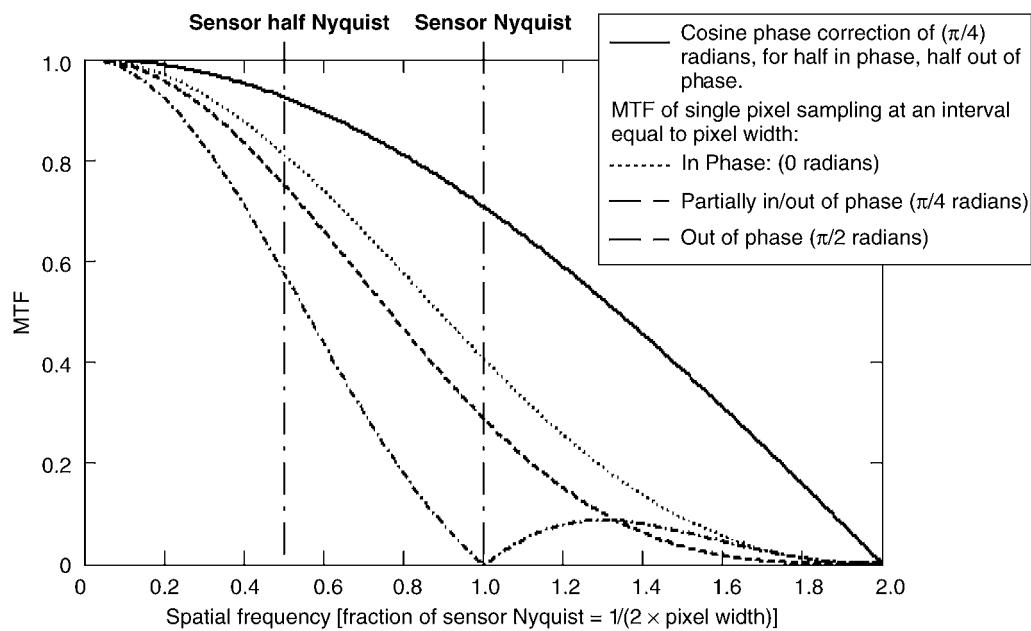


Figure 23.17

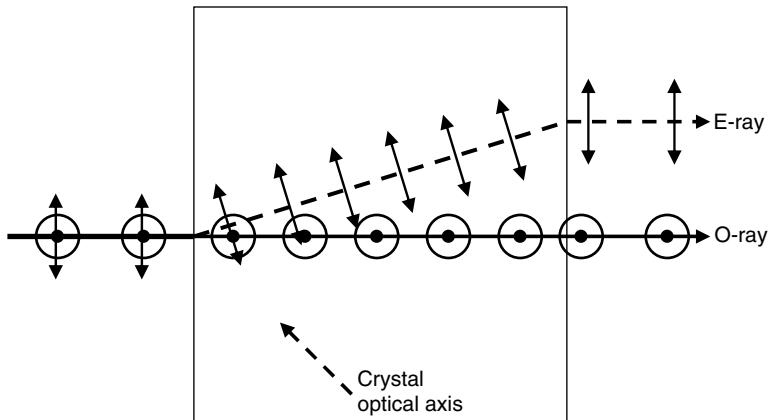
MTF of Pixel and Pixel Sampling Considering Phase

frequency cut-off. With more spot separation, a lower frequency cut-off is obtained. This can be beneficial in improving the overall image appearance, but this can also limit the resolution capability of the optical system, so care must be exercised in its use.

BIREFRINGENT MATERIAL IN AN OLPF An OLPF is made from birefringent material layers. Firstly, the basics behind a birefringent material are briefly considered. Figure 23.18 shows ordinary and extraordinary ray splitting within a piece of birefringent material. Note that the surfaces of the cube are not necessarily at an angle to the incoming ray to achieve ray splitting. However, it is required that the optical axis of the birefringent material is at an angle, as shown.

The E-Field components of the o-ray (in and out of page) are perpendicular to the optical axis. The E-field components experience only one refractive index of the birefringent material and hence the o-ray propagates through the material without deviation. On the other hand, the E-field of the e-ray propagates vertically and hence has an E-field

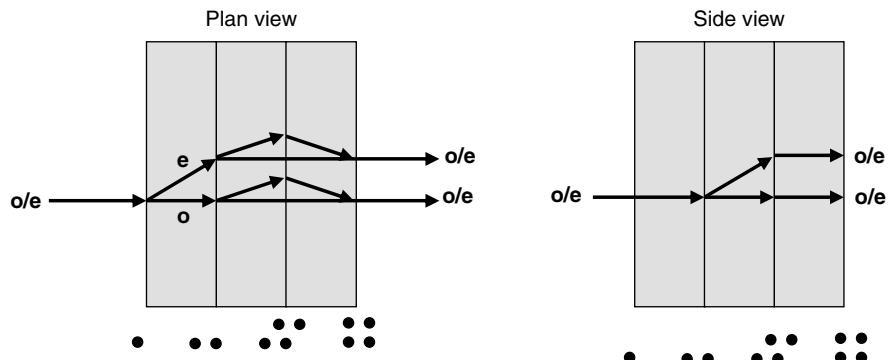
Figure 23.18
Birefringent Ray Splitting



component along the optical axis and perpendicularly to it. As such the e-ray experiences both refractive indices of the birefringent material. The E-field component which propagates along the optical axis does so slower than the perpendicular E-field. This causes an elliptical Huygen's wavelet propagation through the material, which translates the e-beam rather than bending it as occurs in refraction. This offset results in a separation of the o- and e-rays which increases through the thickness of the birefringent material. The o- and e-rays are deviated, but parallel at the output face.

OLPF CONSTRUCTION A simple OLPF can be constructed from three layers of birefringent material. An example is provided in Fig. 23.19. Here we see side and plan illustrations of a filter concept, where the o and e beams are split at various stages, firstly to form a pair of spots, then a

Figure 23.19
Optical Low Pass
(Anti-Aliasing) Filter
Construction



rhombus shape and finally square shape of four spots. The thickness of each birefringent layer, and the orientation of the crystal axis within the each layer of the OLPE dictate the amount of beam separation.

OLPF PERFORMANCE To understand the impact of an OLPE the frequency response of a diffraction limited lens at $f/10$, operating at 500 nm is considered (see Fig. 23.20). An OLPF is then applied to this lens, where the 2×2 spot pattern output by the OLPF is set to 5 μm on a side. The frequency response of this gives a cutoff down to 100 lp/mm, which is half that of the diffraction limited $f/10$ lens. It is noted that some aliasing does still occur; however, this is low contrast compared to the unaliased spatial frequencies.

For comparison, it is possible to achieve a similar spatial filter by halving the $f/\#$ of the system. An $f/20$ lens without an OLPF in place is shown. The advantage of this approach is that there is no aliasing, however, it comes at a high price. We now have an extremely small aperture system giving only a quarter of the light collecting ability.

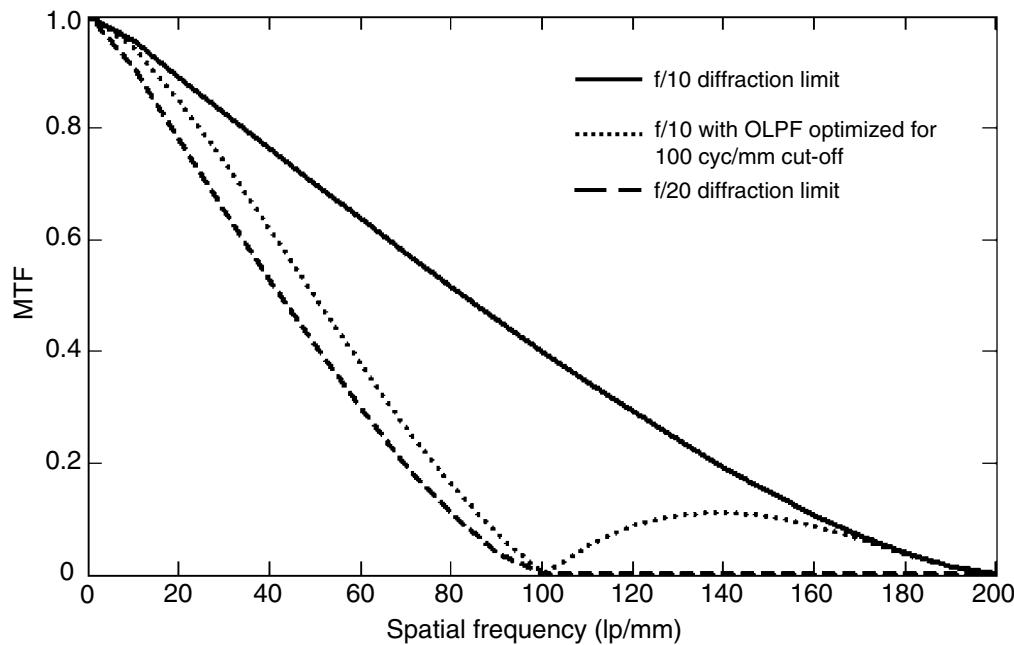


Figure 23.20

Example of impact OLPF on MTF

It is a matter for the engineer to determine if an OLPF is required in a given optical system. In some high-end digital SLR cameras there is an option to slide an OLPF in or out, such is the variability of needs.

A final feature of OLPFs in modern optical systems is that a spectral filter layer is typically built in. Typically in visible systems an infra-red (IR) spectral cut-off filter is employed to limit the detected spectral band to the visible wavelengths. Removing the need to accurately image spectral content which is not required for the application simplifies the optical design requirements, and hence reduces the complexity of the optical design, and in turn the cost. If an OLPF is being used, then why not place the spectral filter on a component which is being installed anyway!

As a final application note on OLPFs: If it is the intent that the optical system is used with a polarizing filter, then care will be required in the design of the polarizer. The OLPF will not operate correctly if linearly polarized light is incident upon it. Photographers often use OLPFs to enhance the deep blueness of sky and to reduce reflected glare from objects (especially from water surfaces). If viewed with an eye, a satisfactory result can be obtained by using a linear polarizer. However, with an OLPF in the optical system, the polarizer would need to be a linear polarizer toward the object followed by a circular polarizer. This allows the unwanted polarization component to be removed from the object scene. The linear polarization state which is desired is then converted to a circular state to ensure that the OLPF will not see a linearly polarized state. Reduced and potentially asymmetric blurring could result giving a higher cut-off frequency (reduced OLPF performance) in one axis compared to the other. This could cause aliasing artifacts to appear in the image. Further information on polarization phenomena can be found in Chap. 19.

Lens Limited Versus Detector Limited Systems

Generally speaking an imaging system comprising a lens and an array sensor would be considered lens limited if the diffraction limited cut-off spatial frequency of the lens [$1/(\lambda \cdot f/\#)$] is less than the Nyquist spatial frequency of the detector. It is usual to design custom optics up to the Nyquist frequency and not to control the optimization for higher spatial frequencies, except in the case where an OLPF is designed in. It may not be possible to achieve a diffraction limited lens design, in which case the lens would have a lower cut-off frequency. The performance across the full field of view required should be considered.

For the optical designer, being detector limited is usually preferable. This is because there is usually less dependency on the lens to yield good overall system performance. In the lens limited case, any variation in performance from one lens to another will directly impact the system performance achieved. Lens to lens variation will certainly occur because of fabrication tolerances. The variation in system performance due to lens performance variability is generally less sensitive in a detector limited system. A lens limited system often occurs either because of competing system and lens design requirements. For example, system packaging and weight restrictions may limit the optical solution space preventing aberrations from being controlled as well as might be desired. A lens-limited situation may also occur if a lens designed for one imaging sensor is matched up to another sensor with smaller pixels, thereby increasing the cutoff frequency of the detector but not that of the lens.

Other Factors Affecting System Level MTF

DETECTOR CHARACTERISTICS Further detector characteristics such as charge diffusion within a detector pixel can be expressed in terms of MTF. Any blurring or smearing effects can be expressed in terms of an MTF.

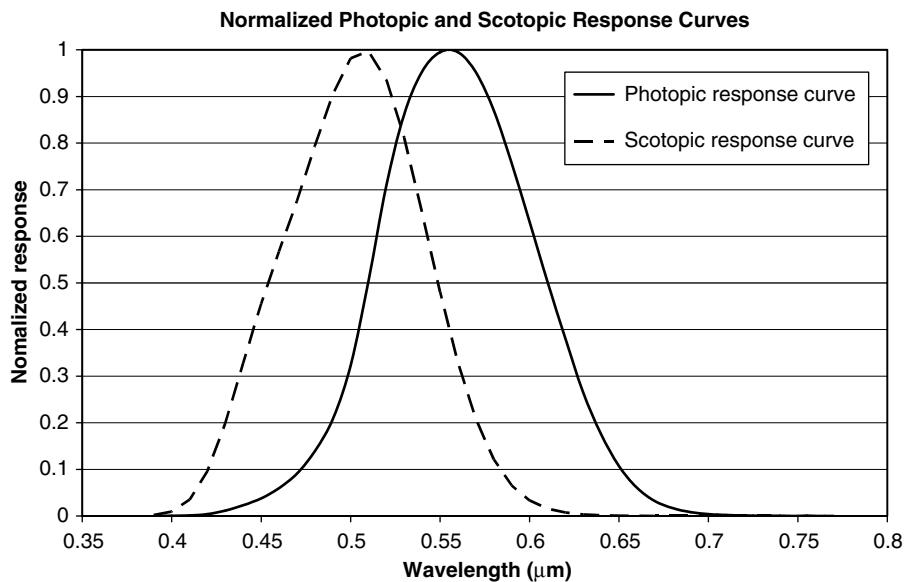
ENVIRONMENT Topics such as atmospheric turbulence, vibration/jitter, stray light, environmental erosion (for example, of windows) can all have MTF factors assigned to them that multiply the lens, detector, target, and sampling MTF plots to provide an overall system model of performance.

HUMAN FACTORS Human interaction with display systems involves visual perception, eye visual acuity, display performances, etc. Eye visual acuity is typically 1 minute of arc and can be used to estimate if normal 20/20 human vision is limited.

One of the most common human factors which need to be considered in an optical design is the photopic spectral response. The dark vision, scotopic response of the eye shifts substantially toward the blue end. A comparison of photopic and scotopic spectral responses is provided in Fig. 23.21. These have been normalized to provide a relative comparison, however, the relative scaling factors for luminous efficacy are 673 lm/W peaking at 555 nm for the photopic response and 1725 lm/W peaking at 510 nm for the scotopic response.

Figure 23.21

Eye Photopic and
Scotopic Response
Curves



Color Sensors

There are three main color imaging architecture types available. These are color filtering at the array level, color sensing within a pixel, or color splitting. These will be considered in turn.

Bayer Pattern Detectors

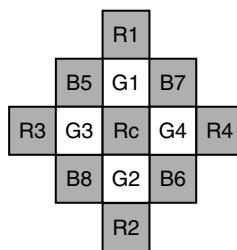
On visible spectral band imaging array sensors a color filter can be applied immediately on top of the CCD array. This is often applied in the form of a Bayer pattern (so called after the inventor Dr. Bryce E. Bayer of Eastman Kodak) (see Fig. 23.22a). Here an RGB pattern is found, where on each row alternating green and red are found on one row and on the next, alternating blue and green are found. This means that there are twice as many green pixels overall. The reason for this is that the eye's response (the photopic response) peaks in the green (Fig. 23.22b). Consideration of the photopic curve tells us that sensitivity in the green is approximately 70% of the total information content, compared to about 20% in the red and about 10% in the blue. Resolution is therefore dominated by the green response. It is further noted that the eye focuses

Figure 23.22

(a) Bayer Pattern
 (b) Green and Blue Interpolation on Red Pixel

G	R	G	R	G	R	G	R	G	R
B	G	B	G	B	G	B	G	B	G
G	R	G	R	G	R	G	R	G	R
B	G	B	G	B	G	B	G	B	G
G	R	G	R	G	R	G	R	G	R
B	G	B	G	B	G	B	G	B	G

(a)



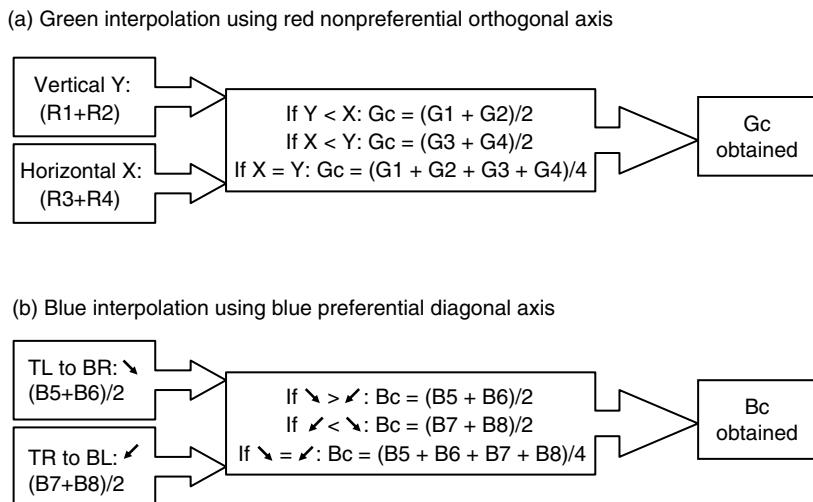
(b)

less well on the blue and red than it does in the green. The need for good green imagery is therefore more important than red or blue imagery.

The Bayer pattern is an efficient manner to obtain color information and retain the full resolution capability of a sensor whilst only using a single sensor. The disadvantage of this technique is that where a blue or red pixel records the image, no green information is recorded. The green content of the image therefore has to be estimated from the surrounding pixels. This technique is known as demosaicing.

DEMOASICING A basic scheme for demosaicing is outlined below. Consider Fig. 23.22b which shows the nearest green to red and blue neighbors. Firstly we want to interpolate the amount of green present in the central red (R_c) pixel. Simple linear interpolation is used. The amount of red found in the horizontal and vertical axes is compared. The axis with the most red content is disregarded. The average of the nearest green neighbors to R_c in the remaining axis is taken to compute the amount of green present on R_c . The computational flow is given in Figs. 23.23a and b. The computational flow for the blue value interpolation on pixel R_c has been computed by using the nearest diagonal neighbor blue pixels. The highest mean diagonal axis is used to compute

Figure 23.23 a and b
Computational Flow
for Linear Demosaicing
of Bayer Pattern



the blue value; in the case where all four pixels are equal, an average of all four is used. Because the blue resolution is typically lower, and because the eye is less sensitive to the blue, it is often reasonable to simply use the average of all four nearest blue neighbors to infer the blue value on the red pixel.

In many practical applications, especially those involving moving video, it is necessary to keep the computation load to a minimum in order to retain high frame rates. However, more sophisticated demosaicing schemes can be employed which use nonlinear weighting, and in movie systems the opportunity exists to use time adaptive techniques. These are beyond the current scope of discussion.

Clearly the value interpolation techniques described are prone to error. Essentially the techniques are estimates which become increasingly prone to error as the Nyquist frequency is approached. For gray scale resolution, where color is not considered, the Nyquist limit is described as $f_{\text{Nyq}} = 1/(2p)$, where p is the separation distance from one pixel to the adjacent pixel (that is, the pixel pitch). It should be noted, however, that the Nyquist frequency for a Bayer pattern in the red and the blue is approximately half of the green, that is, at $f_{\text{NyqBlue\&Red}} = 1/(4p)$. Whilst the green pixels on any given row (or column) are also separated by 2 pixels from center to center, adjacent rows (or columns) are offset by 1 pixel; it is therefore pessimistic to consider the green Nyquist as being half the grayscale case and in fact reasonable to consider it still as $f_{\text{NyqGreen}} = 1/(2p)$.

The implication of this is that the frequency response of the green is better than that of the blue and red. Color artifacts can, therefore, occur in the image. Whilst the green spatial frequencies may be correctly detected up to the sensor Nyquist frequency of $1/2p$, incorrect frequency responses may occur at lower frequencies (at $\sim 1/4p$) for the red and the blue. For example, it is possible to imagine a blue “picket fence” of vertical bars at the green Nyquist spatial frequency of $(1/2p)$, where the blue bars fall onto red and green pixels columns and the clear gaps fall on the blue green pixel columns. In this case, the blue signal will not be detected at all. In a second scenario, if the blue vertical bars align with the blue and green pixel columns, and the gaps align with the red and green pixel columns, then every blue pixel will detect a bright blue signal. A solid blue color (with no modulation) would be interpolated, that is, the blue $1/(2 \text{ pixel})$ Nyquist modulation is aliased as a blue continuous bias across the image. There are clearly an infinite number of cases in between these examples where the blue picket fence is out of phase with the sensor pixels where careful computation is required to understand the nature of the detected signal.

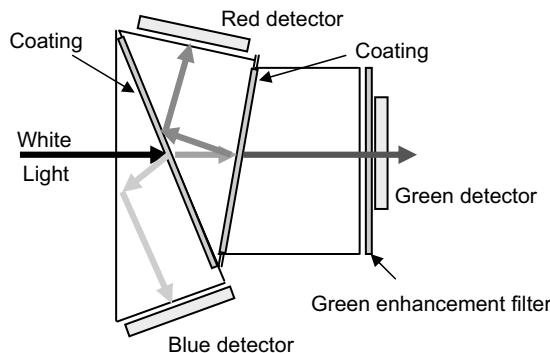
Three Color Sensor

To avoid the color artifact issues which can arise with Bayer patterns, and the computational load incurred by demosaicing the Bayer pattern, a three focal plane architecture can be adopted.

COLOR SPLITTING PRISMS Figure 23.24 shows a three sensor architecture where each focal plane is dedicated to a single color. In this prism architecture the optical path from the 1st optical surface to each focal plane is usually identical. However, it is noted, that it does not have to be. If for example, axial color was proving to be a difficult aberration to control, then each focal plane could be a slightly different length. This may help to further optimize the lens performance, however, care needs to be taken that each image does not scale differently.

A potential disadvantage of this scheme is that each camera must be aligned to the others in the horizontal and vertical positional and angle to achieve the best performance. Furthermore, this alignment must be maintained throughout the life of the optical system. The use of a relatively large beam splitting optics close to the focal plane requires that the design have a long back focal length.

Figure 23.24
Three Camera Array
Color Splitter Prism
Arrangement



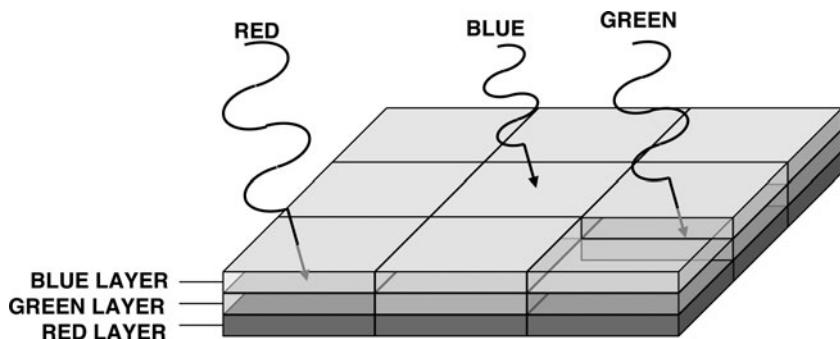
It is worth noting that in LCD projector display systems an alternative to the trichroic prism is the X-prism cube, where it is employed as a beam combiner. Red, green, and blue images are combined into a single composite RGB image using the X-prism.

In applications where color splitting is used, if each of the three arrays is aligned pixel for pixel to the object, then the original resolution of the array can be maintained. Some camera systems deliberately offset the red, green, and blue CCDs by a fraction (say half) of a pixel. In this way, subsampling of the object can be achieved. From this it is possible to interpolate higher resolution than any single sensor would be capable of. Of course, this takes careful alignment of each CCD and complex algorithms to infer the enhanced resolution.

COLOR SENSING WITHIN PIXELS Layered sensors have been developed to allow the detection of different wavelengths on one pixel. An example of this is the Foveon sensor. A schematic of this is provided in Fig. 23.25. Here we observed that the blue information is absorbed before the green, and finally the red information is detected. The spectral sensitivity of each pixel will be subject to manufacturing and electrical variations. As such each pixel's spectral response has to be calibrated and normalized. Cross talk and loss of signal due to reflections can be an issue.

A great advantage of this technology is that full resolution is achieved in all three color bands. Furthermore, any ruggedness issues associated with camera to camera alignment are eliminated. The package volume can also be substantially lighter than a three-sensor package due to that lack of need for beam splitting optics.

Figure 23.25
Foveon Sensor



Electronic Correction

A growing area of influence on the optical design is the role of electronic correction. Examples of electronic correction have become increasingly evident in electronic imaging systems with the rise in popularity of the staring array. Many cameras include post-sensor frame capture correction, but there is a rise in the use of on-chip processing often inspired by nature—so called biomorphic sensors. The advantage of on-chip processing is that it enables faster processing. This, however, often reduces flexibility.

Early examples of electronic correction were found in thermal imaging systems to correct residual Narcissus effects which create a non-uniform appearance in the image. A background image of a uniform scene is stored electronically and then subtracted from all subsequent frames allowing a uniform appearance to be recovered.

The same uniformity correction concept is found in modern day commercial visible band digital cameras, where the vignetting varies across the field in addition to the $\cos^4(\theta)$ roll off which occurs naturally. Variation in illumination across the image plane is often referred to as relative illumination. The center (normally the maximum) is given a value of 100% with points elsewhere “rolling off” to lower values. Typically such cameras will have zoom lenses. As the zoom is varied, the amount of relative illumination across the sensor plane changes. A lookup table is used which contains different uniformity correction tables for each zoom position.

Distortion correction to remove pin-cushion or barrel distortion can also be applied. This is usually done in remote software packages. Care has to be taken in the amount of distortion correction which can be

achieved. Aliasing artifacts may occur in the image if too large an amount of image correction is attempted.

Lateral color correction is becoming increasingly popular in projection applications using narrow band laser or even LEDs. By storing different distortion maps for each spectral band, it is possible to free up lateral color as a design constraint. This often simplifies the glass selection and overall design complexity.

Overall, electronic correction can be of huge benefit to the optical engineer, if correctly employed. Designing with electronic correction freedom usually relieves the pressures on the optical design and allows design choices to be made which may not have been possible before. A simple example would be to avoid using an aspheric element to correct distortion in a design. Of course, designing in electronic correction in a system means that the electronic correction is required to get to an operating system. This has to be remembered when evaluating the optical subassembly prior to integration with the full optical system.

Camera Connectivity

In any optical system, the electrical interface with the sensor must be considered. This is an important consideration for the optical systems engineer at the outset of the optical system design. Important considerations include the rate at which data can be acquired from the sensor array, the robustness of the connectivity in terms of frame latency and reliability, that is, likelihood of dropped frames occurring.

Several different architectures have been developed which have different strengths and weaknesses. A comparison of commonly available camera connectivity types is offered which considers performance and relative costs. A summary is provided in Table 23.3.

Camera Link

Camera Link offers a robust, high data rate image acquisition solution. This offers the bench-mark performance for high fidelity, high data-rate environments. Camera Link offers high data rates (up to 5.44 Gbps) and high robustness to lost frames. Camera Link is therefore a favored connectivity choice for high speed in-line inspection applications such as assembly line bar code inspection.

TABLE 23.3

Summary of Camera Connection Types

Camera Connection	Type	Bandwidth (Mbps)	Scalability	Cable			Frame Grabber Required?	Computer Interface/Connector	Latency
				No Switch	Switch	Fiber			
Camera Link : base specification	2040	continuous	1	10 m	n/a	n/a	Y		Low
Camera Link : medium specification	4080	continuous	1	10 m	n/a	n/a	Y	PCI Framegrabber	Low
Camera Link : full specification	5440	continuous	1	10 m	n/a	n/a	Y		Low
USB	12 (USB1.0)	burst	127	<5 m	30 m	n/a	N	PCI or Motherboard	High
	480 (USB 2.0)	burst							Medium
FireWire	800 Mbps, but only 540 Mbps for image data	continuous	63	<4.5 m	72 m	200 m	N	PCI or Motherboard	Medium/low
GigE Vision	~1000	continuous	Unlimited					GigE NIC or LOM (LAN on motherboard) link with RJ-45/Cat-5.	High
				<100 m	Unlimited	Unlimited	N		
10GigE	~10 Gbps	continuous	Unlimited						High

This connectivity architecture requires a camera equipped with the Camera Link interface and a frame-grabber card in the image acquisition computer. Camera Link connectivity provides data transmission for the image, camera control data and asynchronous serial communications on one cable. Power must also be supplied to the camera and acquisition computer (including frame-grabber card). There are three levels of Camera Link specification referred to as basic, medium, and full specification implementations. The medium and full specifications are essentially replicated versions of the basic implementation offering identical camera communication and control functionality but with increased image acquisition data rates. The ability to send camera control data allows Camera Link to be reconfigured without affecting the image acquisition data flow. This ability may be needed in a rapidly changing environment such as a critical moving part inspection environment.

The Camera Link standard is based on the Channel Link LVDS (low voltage differential signaling) chip set manufactured by National Semiconductor. The logic operates at 3.3 V, hence the low voltage. The Channel Link interface consists of four data transmit and receive channels and a single clock transmit and receive channel. 28-bit data is handled in CMOS TTL (Transistor-Transistor-Logic) circuitry. The basic implementation transmits 28-bits at 595 Mbps, therefore 2.38 Gbps (297.5 Mbps). Of this, image data up to 24 bits can be transferred. This gives an image bandwidth of 2.02 Gbps. The medium and full implementations offer 48 and 64 bits image data transfer and give data rates of 4.08 Gbps and 5.44 Gbps respectively.

Disadvantages of Camera Link are that it does require a frame-grabber card to work which adds to overall system cost. It also does not allow interconnectivity of several cameras to one controlling computer. It only allows one camera at a time to be connected. The high data rates are in part achieved by limiting the data cable length to 10 m. Camera Link is therefore not suitable for applications where the camera has to be located remotely from the data computer.

Table 23.3 summarizes the performance of Camera Link in each of the three specification levels, base, medium, and full.

IEEE1394b

IEEE1394 is more commonly referred to as Firewire, which is Apple Inc's proprietary name for this interface. The IEEE1394 standard covers the

commonly used IEEE1394a at 400 Mbps and IEEE1394b at 800 Mbps, but also provides provision for future bandwidth increases up to 3.2 Gbps with appropriate cabling. Of the 800 Mbps in IEEE 1394b, 540 Mbps is available for image data transfer.

Firewire is not limited in extensibility as Camera Link is, but does not offer the unlimited capability of GigE. Up to 64 cameras can be interconnected but simultaneous connection to a single PC would be limited by the PC bus speed.

Other benefits provided by Firewire are low numbers of dropped frames and a frame grabber card is not required. The standards are being developed to be backward compatible. With future bandwidth growth through to 3.2 Gbps, there is significant improvement potential for upgrading without redesigning. Cabling can run to ~72 m, which is long enough for most factory based applications.

USB

Similarly to Firewire, USB doesn't require a frame grabber. USB is found on most new current computers, so is an obvious choice for compatibility. The bandwidth rate of 480 Mbps is reasonable. Whilst this may never be a preferred choice for scientific applications or for high speed/data critical applications where lost frames are not acceptable, it does offer a low cost alternative.

Gigabit Ethernet (GigE) Vision

Gigabit Ethernet Vision, commonly referred to as GigE Vision, or just GigE, is a standard which defines how to configure a camera for connection to a central host computer using standard ethernet networking and interfacing. It operates by assigning each camera on a network an IP address, thereby allowing each camera to be uniquely identified. Data channel streaming, the transmission of camera control and image data are also defined by the GigE Vision standard.

Existing ethernet networks can be used to connect cameras together. This clearly offers a great deal of flexibility in the implementation. However, the chain is only ever as strong as the weakest link! If a component along the network from the camera to acquisition pc is limited by a single 100 Mbps component, the entire acquisition rate drops from

1000 Mbps (1 Gbps) down to 100 Mbps. Care therefore has to be taken to ensure that the entire network can deliver the required bandwidth.

A great strength of GigE is the ability to connect many cameras together on one network. A single controlling computer can be used to access multiple remote cameras. This offers unlimited camera connection scalability. Without a network switch in place, connections can be up to 100 m. Once network switches are put in place, the communication distance is unlimited.

A disadvantage is that the protocol does suffer from image latency. Increases in latency are to some extent inherent in the image and control data packing requirements of the GigE standard. Cable length does add to latency to some extent, much in the same way that internet communication rates can slow with distance. Periodic dropped frames may occur. GigE is therefore not best suited to high speed process control applications where low latency is required.

GigE Vision offers an excellent choice for applications which are remote, do not require low latency, and are not critically hampered by the occasional lost image frame. Surveillance is an example ideally suited to GigE, where several remote cameras are connected to a single central computer, and high speed frame rates are typically not required.

Future developments of the GigE Vision standard toward 10 GigE will offer data rates up to 10 Gbps. This improved standard has been initiated but has not been finalized at the time of writing. Of course, existing networks will only benefit from a 10 GigE standard if all components in the network chain deliver a 10 Gbps bandwidth.

Bibliography

- Bayer, B.E. *Color Imaging Array*. Patent US3,971,065. US Patent & Trademark Office—uspto.gov.
- Egri, J. *Test and Measurement World* (www.tmworld.com)—Ethernet vs CameraLink. Imperx: May 2006.
- ERIM/SPIE The *Infrared & Electro-Optical Systems Handbook*, Vol 6: ISBN 0-8194-1072-1.
- Fischer, R.E., Tadic-Galeb, B. *Optical System Design*, 1st ed. McGraw-Hill: 2000; ISBN 0-07-134916-2.
- Hecht, E. *Optics* 2nd ed. Addison Wesley: 1987; ISBN 0-201-11611-1.

- Holst, G.C. *Electro-Optical Imaging System Performance*. 1995; ISBN 0-9640000-1-6.
- Janesick, J.R. "Dueling detectors," *OE Magazine* Feb 2002.
- Janesick, J.R. et al., "Using shot noise to measure sensors: Charge-coupled-device charge-collection-efficiency and the photon-transfer technique," *Optical Engineering* Oct 1987; 26(10): 972–980.
- Lyons, R.G. *Understanding Digital Signal Processing*, 2nd ed. Prentice Hall: 2004; ISBN 0-13-108989-7.
- RCA Electro-Optics Handbook*. 1974; EOH-11.
- Senior, J.M. *Optical Fiber Communications—Principles and Practice*, 2nd ed. Prentice Hall International Series in Optoelectronics: 1992; ISBN 0-13-635426-2.
- Wilson J., Hawkes, J.F.B. *Optoelectronics: An Introduction*. Prentice Hall International Series in Optoelectronics: 1989; ISBN 0-13-638495-1.

Stray Light and Optical Scattering

Introduction

Stray light can be the curse of an optical system. The optical engineer may toil to optimize and build a system, only to have the performance destroyed by an unexpected glare, glint, or smear of light! It is often possible (but not always, so proceed cautiously!) to fix the causes of stray light in a system during the debug phase, once hardware is in hand. However, good engineering judgment and practice can be applied to minimize this effort and save resources in the latter phases of a project. This section discusses different causes of stray light in an optical system, how it can be described and managed, and finally how it is even possible to make use of it.

Stray Light Scatter Sources

The key to solving optical scattering problems in the optical system is to identify the source of the problem. In an ideal world, the designer would model all aspects of the system to ensure fully compliant performance is achieved before the build phase is attempted. However, such exhaustive modeling is usually not possible because it can be incredibly time consuming, and results can sometimes be misleading. Great care is required in interpreting the stray light model; this is discussed further in the

modeling section (see “Modeling & Analysis Techniques” section) which follows. In optical systems where stray light is expected to be an issue, there is often some level of empirical work required to ensure that the optical system performance is satisfactory. One should always plan for this debug and evaluation effort in a real optical system build and test.

Bounces and Ghosts

VISIBLE SYSTEMS A good place to start in evaluating an optical system for stray light is to check for double bounces or ghost images. Double bounce stray light reflection occurs when light is reflected from one optical surface backward and subsequently forward again toward the focal plane. See Figs. 24.1 and 24.2 for an example. The designer should check to ensure that the double bounce does not focus onto the focal plane. In practice the double bounce is seldom a problem in a visible optical imaging system, particularly if the optical elements are coated. However, it is always worth checking to make sure.

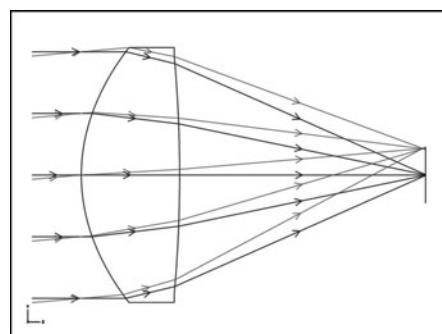
In the visible band, uncoated glass surfaces will provide a surface Fresnel reflection [$R(\%)$] computed from the following equation:

$$R(\%) = 100\% \cdot \left(\frac{n_{\text{GLASS}} - n_{\text{AIR}}}{n_{\text{GLASS}} + n_{\text{AIR}}} \right)^2 \quad (24.1)$$

For $n_{\text{AIR}} = 1$ and $n_{\text{GLASS}} = 1.5$ we have 4% reflection per surface.

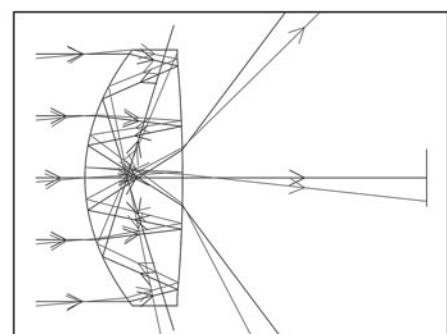
Figure 24.1 and 24.2

Double Bounce Example Using a Singlet



Single imaging two field points.

(24.1)



Double Bounce : Bounce off 2nd surface and then 1st surface does come to a focus, therefore is not a concern in this example.

(24.2)

Even a relatively low cost antireflection (AR) coating (for example, MgF₂ coating) will provide a peak transmission of at least 98%. The 2% reflection double bounce would be only 0.04% of the original energy. This gives a scatter signal to image signal ratio of 1:2500, that is, the scattered return is weak compared to the signal. If the scatter is not focused onto the imaging sensor, then the scatter signal further weakens compared to the image signal. In most cases, double bounce reflections will only show up as a scatter artifact on the image if a bright object is in the scene *and* this causes a double bounce reflection to focus on a dark part of the image.

In Figure 24.3 an example is provided where a commercial grade digital camera images a flashlight against a black background. This is a severe stray light test for an optical system. Whilst this system is not likely to be often used in this mode of operation, it does serve to highlight stray light artifacts which exist within an optical system. A halo around the flashlight is noted. This is caused by optical surface scatter from the lenses within the optical system. The blob of light noted to the left of the flashlight is a double bounce scatter within an element of the lens system; as the flashlight is moved to the right, the blob moves to the left.

Figure 24.3
Commercial Digital Camera Exhibiting Scatter Artifacts in Presence of Very Bright Flashlight Object. Note the Bright Halo and Scatter Spot

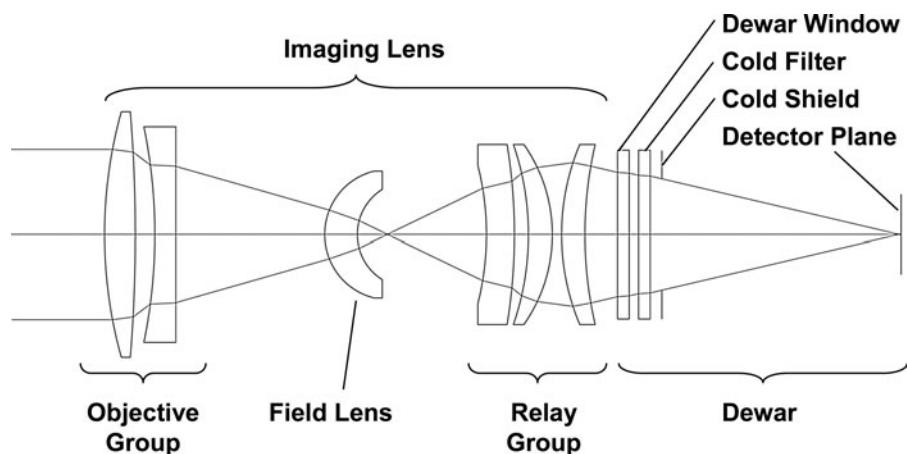


INFRARED SYSTEM CONSIDERATIONS AND NARCISSUS In infrared (IR) systems higher refractive index materials are typically used. Typical examples are Silicon ($n_{\text{Silicon}} \approx 3.42$ at $4 \mu\text{m}$) and Germanium ($n_{\text{Germanium}} \approx 4.02$ at $4 \mu\text{m}$). Uncoated in air, Silicon and Germanium have significantly larger Fresnel reflections at IR wavelengths than glass surfaces do in visible wavebands. A BK-7 glass surface in air exhibits approximately a 4% reflection in the visible waveband, compared to a 30 to 36% reflection for Silicon and Germanium, respectively. AR coating of IR materials is therefore imperative not only for ghost bounce scatter suppression but also to ensure good optical transmission through the IR lens system.

IR imagers offer a unique form of optical scatter which is worthy of special consideration. Most IR imagers operating in the 3 to $5 \mu\text{m}$ mid wave IR (MWIR) and 8 to $14 \mu\text{m}$ long wave IR (LWIR) bands use sensors which are cooled. When operating in these spectral bands objects the designer must remember that surfaces at room temperature are blackbody emitters; that includes the optical materials and the lens housing walls.

An example of a typical IR imaging system is provided in Fig. 24.4. Features of the imaging lens include a color corrected doublet lens (of Silicon and Germanium) in intermediate focus which is preceded by a field lens for aberration correction, and a relay group which reimages the intermediate focus onto the focal plane. The camera includes a Silicon entrance window, a cold filter, a cold shield (stop) and the detector plane. The components following the window are all maintained at a cold temperature.

Figure 24.4
Example of Infrared
Lens Design

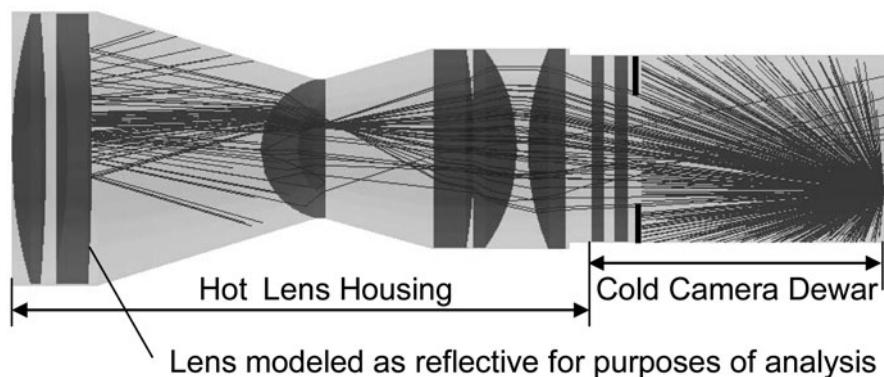


The intermediate focus is required to achieve a 100% cold shield efficient design. 100% cold shield efficiency means that any point on the detector plane which exits through the cold shield aperture will always pass through a lens and out into the scene. There is no direct path from the focal plane, through the cold shield and onto the hot housing wall.

In practice, there are several factors which cause thermal contributions to be coupled into the optical path, even in a 100% cold shield efficient nominal design. Fresnel reflections are the most significant contributors, especially if the optical elements are not properly coated. Other factors include contaminants on the optical surfaces, and the emissivity of the lens materials themselves (albeit that this is a low and unavoidable contribution), so minimizing the number of optical elements should always be part of the optical designers goal. Care should be taken of the optical coating performance with angle to ensure that it does not significantly fall off in transmission over the range of angles of incidence which each surface experiences.

Consider Fig. 24.5. There are several things to note on this figure. Firstly, a point has been set up as an imaginary source on the edge of the focal plane. This allows us to consider what this edge pixel on the detector will actually see. Firstly, we note that many possible ray paths terminate within the cold camera dewar. The dewar is cold, so no thermal signature is contributed by the ray paths which terminate on the dewar walls (or vice versa, start on the dewar walls and end on the detector). Now, again trace backward from the detector and consider the ray paths which exit through the cold shield. We see that these ray paths continue, without reaching the housing walls directly. Eventually we reach the second objective lens. This has been modeled

Figure 24.5
Narcissus—Examples of “Hot” Housing and “Cold” Detector



as a reflective element to examine the effect of the Fresnel reflection. Ray paths from this point are indirect paths. Now, tracing forward, we can see that it is possible for a ray which starts on the hot housing wall to undergo a Fresnel reflection and then reach the cold detector. This means that a hot housing wall will contribute some thermal signature to the detector. This contribution can be minimized by good optical AR coatings.

The background thermal signature obtained at any given point on the field is the sum of all of the “hot” housing and “cold” detector/dewar contributions. If every point on the detector plane experiences the same hot and cold sum, then a uniform offset would occur across the field. This can effectively be removed by applying an offset to the image.

Problems occur when there is a difference between the background thermal contributions across the focal plane. Imagine a point in the center of the detector, for which a Fresnel reflection focuses back down onto the detector; this point will experience a cold return. If at the same time a point at the edge of the detector experiences Fresnel reflections which do not hit the cold detector, but instead see hot wall structure, then a *hot* contribution will be experienced. In an image where white is displayed as hot, we will see a black hole in the center of an image which has not been uniformly corrected. This effect is known as *Narcissus*, named after the tragic figure from Greek mythology that fell in love with his own reflection. If your lens design ends up with Narcissus, it might well be a tragic sequel!

Nonuniformity correction (NUC) allows a background calibration image to be subtracted from all subsequent images; this is recorded when a uniform object scene is present. This allows the background effect to be removed, so long as the Narcissus profile across the detector field does not vary. Unfortunately, because the *hot* walls are often not in a thermally stabilized environment, the hot contribution to the edges of the pixels will vary, whilst the cold contribution at the center of the image does not because it is in the thermally stable environment of the cold dewar. If the hot walls now cool down, and the original uniform calibration data is applied, a hot center will be experienced, with comparatively cold image edges. In IR systems experiencing a wide thermal environment it is often necessary to acquire calibration data at several temperatures and to apply the closest NUC table, or otherwise interpolate between NUC sample points to ensure an appropriate uniformity correction is applied across the field. It should be noted also that the emissivity characteristics of the housing surface will also impact the

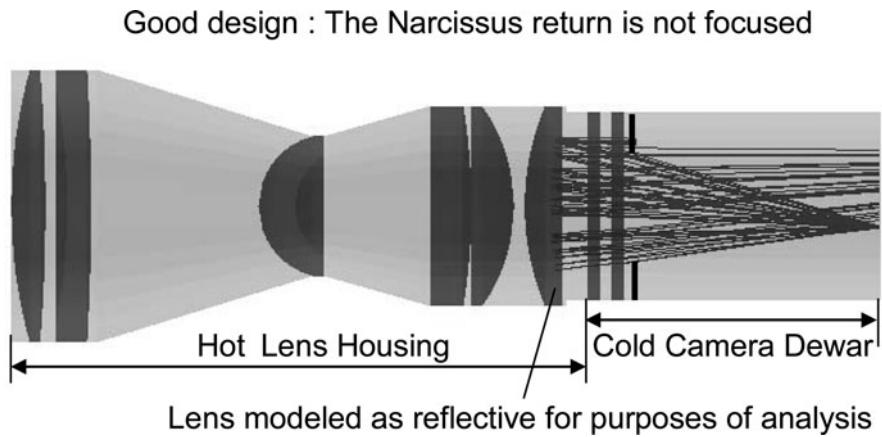
thermal contribution and may cause nonlinear variations in the overall Narcissus signal with environmental temperature.

As a final word, avoid the temptation of fixing everything electronically after the fact. Whilst electronic uniformity correction will aid the optical design, any corrections which are applied will eat up the available dynamic range detector. Good Narcissus design is, therefore, imperative if optimal performance is to be achieved from the IR optical system.

How to Minimize Narcissus in the Optical Design Narcissus is a special case and is in effect a single bounce consideration. A Narcissus return is unavoidable on axis. However, Narcissus can be effectively eliminated by spreading the cold return evenly across the detector. Figure 24.6 shows an example of an axis bundle which does not generate a focused Narcissus return. Whilst the elements in the eyepiece closest to the image plane are normally the elements to take the most care over, it is essential that Narcissus generated by Fresnel reflections be examined for all optical surfaces in the system.

Many optical design packages are now flexible enough to allow a nonsequential model to be generated where the housing walls and optical elements and the detector can be treated as optical sources. By examining the uniformity profiles obtained on the detector, it is possible to model the stray light paths which will cause a Narcissus effect.

Figure 24.6
Anti-Narcissus Design



Structural Scatter

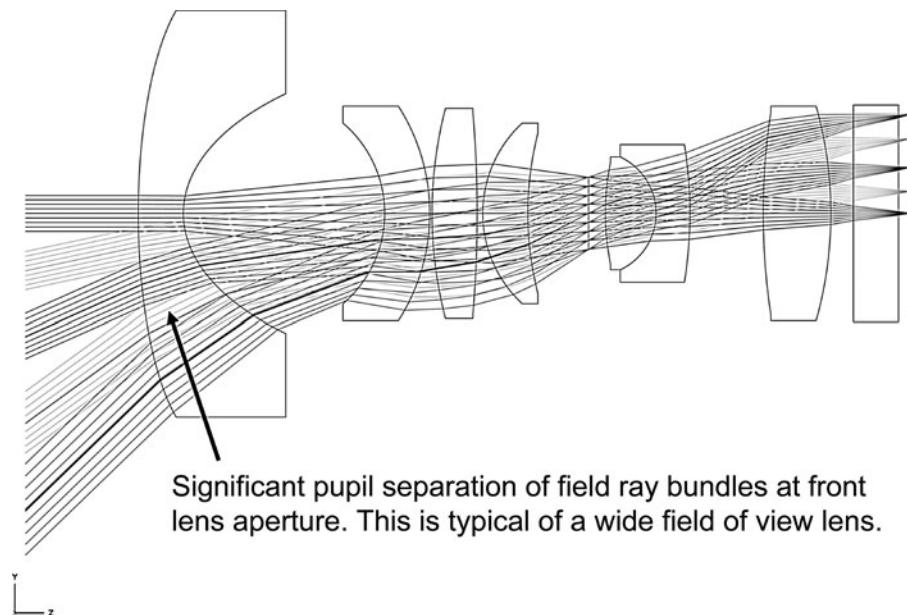
Structural scatter can be more difficult to cope with. This is where the mechanical housing of the lens scatters unwanted light back into the optical path. The source causing the scatter may not even be in the geometrical path of the optical design.

Structural scatter can be minimized using the following techniques:

- Rifling and blacking the lens inner barrel
- Using strategically placed baffles
- Applying blackening to the edges of the lenses
- Minimizing the pupil positional variation across the aperture at the first lens surface; this is particularly an issue in wider field of view lenses

Structural scatter is often observed in wide field of view systems. This is because the front lens aperture diameter is typically large with respect to the size of the field ray bundles entering the lens. See Fig. 24.7 for a typical example of a wide field of view lens which

Figure 24.7
Wide Field of View
Lenses Typically
Exhibit Significant
Pupil Separation at
the Front Aperture



exhibits a large amount of pupil separation at the front element. Such a large aperture increases the possibility of stray light to be coupled into the system.

Whilst the stray light may have no direct path to the focal plane, there is an increased probability of light entering the bore and experiencing structural scatter. If the front aperture is minimized, this can help in reducing the potential scatter pathways. The ideal solution from a stray light perspective is to have the stop reimaged at the front aperture. However, this may serve to over-constrain or over-complicate the optical design. Attempting to minimize the extent of the aperture will, however, help. Note that we are attempting to minimize the aperture diameter by controlling pupil separation, and not by reducing the entrance pupil diameter. Changing the entrance pupil diameter would slow the system thereby changing the fundamental optical performance of the lens.

Low Light Level Systems

Stray light becomes an acute problem when the strength of the stray light is large with respect to the object or signal being observed. This often occurs in systems where light levels are low and correspondingly the sensitivity of detectors is designed to be high and exposure times are increased.

Astronomical telescope systems are a good example of where detailed stray light modeling is required. Moon and star light hitting the telescope mirror support can lead to severe stray light problems. Astronomical telescopes use highly sensitive detectors, and often use long exposure times. Under these conditions even very low levels of optical scatter may be large compared to the signal strengths being detected. Often approach roads to telescopes require that lights are dimmed or turned off near to the facility to prevent accidental stray light from headlights entering the telescope structure.

Types of Scatter

Linear optical scatter can be characterized in two main groups, Rayleigh and Mie. These are described below. Linearity refers to linearity with respect to optical flux, that is, more light will give proportionally more

scatter. Nonlinear optical scattering phenomena such as Raman and Brillouin scattering are not discussed here. These forms of scattering are often significant in laser and fiber-optical communication applications.

Rayleigh Scattering

Light scattering caused by objects which are very small compared to the wavelength of light is known as Rayleigh scattering; aerosol gases are a good example. Lord Rayleigh defined scattering in terms of the following relationship:

$$\text{Rayleigh Scatter} \propto 1/\lambda^4 \quad (24.2)$$

This inverse forth power dependency on wavelength explains why blue light is scattered more than red light. Consider white light incident on the Earth's atmosphere. Blue light is scattered more by the atmosphere than red light. From Eq. 24.2 we find that blue light at 450 nm scatters 4.4 times more than red light at 650 nm. Without the atmosphere, the sky would appear black during the daytime, just as was observed by astronauts on the surface of the moon. On Earth, the blue component of white light is scattered across the atmosphere, which is why the sky is blue! Furthermore, as the sun falls low on the horizon, the amount of atmosphere through which the sun's rays pass increases. This increases the amount of scattering, especially of the shorter wavelengths. With a greater loss of shorter wavelengths, the observer's perception is that the sun becomes more red; in actuality it is becoming less blue. This explains why the sun turns increasingly orange and through to red during a sunset.

Mie Scattering

Light scattering caused by objects which are large compared to the wavelength of light is known as Mie scattering. Generally speaking, Mie scattering becomes significant when the scattering particle is larger than approximately 1/10th of the wavelength of light.

Mie scattering theory is a general theory for scattering caused by any diameter of spherical particle. Mie scattering is particularly useful for colloids. This is a key form of scattering in optical fiber communications, where scattering must be minimized along long lengths of optical fiber to achieve good communication distances.

Modeling and Analysis Techniques

Ghost Analysis

Most optical ray tracing software packages provide some form of ghost analysis capability. In sequential optical models, a common technique often applied to the examination of double bounce ghosts, is to:

- Turn a refractive optical surface into mirror, to generate the first bounce
- Duplicate the preceding optical surfaces
- Turn a second optical surface into a mirror, to generate the second optical bounce
- Trace back to the focal plane

The number of combinations of single bounces increases with the number of refractive surface in the lens. Assuming n refractive surfaces, the number of double bounce combinations (DBCs) is given by:

$$\text{DBCs} = \sum_1^n (n - 1) \quad (24.3)$$

In each case the convergence of the rays on the focal plane is examined. If the spot is too small (typically a double bounce spot size of <1 mm diameter is too small), then the curvatures of the offending pair of double bounce surfaces must be modified to increase the double bounce spot size, thereby reducing the double bounce ghost problem.

In a single bounce analysis each surface is in turn converted to a mirror, and the position where the ray terminates is examined. A single bounce analysis is often used in systems where there is an illumination source and a detector in a separate optical channel.

The single bounce analysis can also be used for Narcissus analysis where the cold detector is converted to an optical source. In this way the cold signal can be designed to be defocused on the focal plane thereby reducing the Narcissus return. An alternative approach to Narcissus analysis is to turn each object into an emitter, including the housing walls; some software packages offer sufficient flexibility in the source shape definitions to make this possible. This approach allows each object to be modeled as a blackbody emitter, and can therefore be a more accurate approach.

Scatter Path Analysis

To ensure that structural scatter is not a problem, sample object field points are typically set up at a representative range of angles. An example of this is provided in Fig. 24.8.

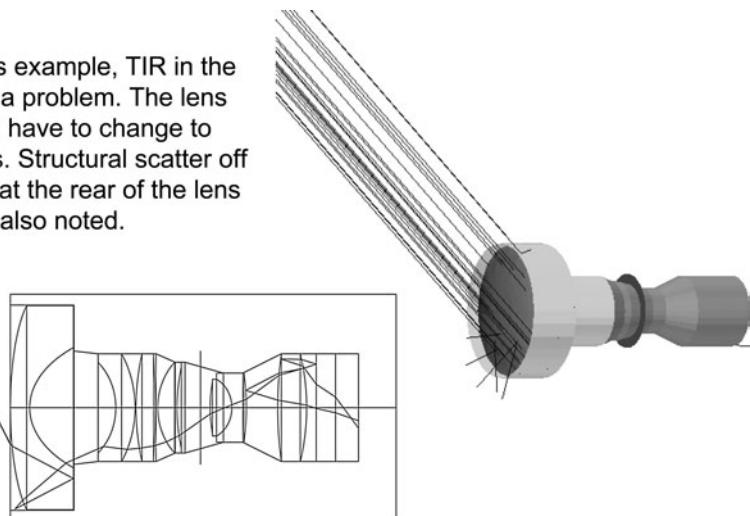
Figure 24.8 highlights an optical path through the system which must be prevented. This can be achieved either by the addition of a lens hood or a structural baffle within the bore. In the example shown in the figure, the walls are modeled as mirrors or as highly specular surfaces. This allows the worst case ray paths to be identified. Further refinement is offered by modeling the walls as scattering objects. Eliminating single specular reflections and single diffuse reflections will suppress most scatter in most optical systems. Refining the analysis further to consider secondary scatter bounces may be required in scatter sensitive applications.

The sample field points selected should extend out of the geometrical field of view of the lens design. A bright source such as a lamp or the sun may still be scattered onto the focal plane via the structure.

Doublet and triplet lenses are worthy of special consideration when conducting a stray light analysis. In some cases a ray path will be found where total internal reflection (TIR) occurs internally in a lens element. In the case of a doublet or triplet element, the TIR condition

Figure 24.8
Example of Scatter Path Analysis Modeled Using Discrete Field Point Ray Bundle

Note : In this example, TIR in the first lens is a problem. The lens shape will have to change to suppress this. Structural scatter off the housing at the rear of the lens is also noted.



is affected by the optical cement used to bond the lenses together. This layer has to be considered carefully, especially in the case where the lens elements are of particularly high or low refractive index, that is, where the refractive indices of two bonded elements differ significantly from that of the optical adhesive. An optical cement will typically have a refractive index of approximately 1.5, up to 1.6 or so for a higher index optical cement. The seamless high index interface found in the optical model is in reality interspersed with a thin layer of lower refractive index material; a 50-mm-diameter lens may have a layer of approximately 25 μm thick. Whilst this layer has a negligible effect on the direct ray paths through the lens and so does not impact the lens imaging performance, it does change the critical angle at the interface. As such, TIR behavior is modified. In order to accurately trace stray light paths through the system, the doublet and triplet lenses need to be accurately modeled to include the optical cement layer. It is normally the easiest to conduct this analysis in a nonsequential stray light model.

Modeling Scatter—Watch The Magnitude!

When examining stray light sources it is important to estimate the relative magnitude of the stray light signal strength with respect to the image. In order to examine optical scatter within a model may require increasingly small sources to be set up. It is easy to lose track of the magnitude of the signal which actually reaches the source. It is a good practice to compute the scattered light in terms of the total scene energy. This provides a basis for normalizing the scattered energy which reaches the detector.

Veiling Glare

This is a form of optical surface scatter which occurs when object field point rays and rays from objects outside the geometrical field of view of the optical system strike a lens surface and are thereby scattered into the optical system. Veiling glare is generally low in structure and is, therefore, uniformly scattered over the focal surface. Despite the low structure of the glare, a reduction in image quality is observed. This

can be explained by considering the contrast equation (Eq. 24.4), then we have:

$$\text{Contrast} = \frac{I_{\text{MAX}} - I_{\text{MIN}}}{I_{\text{MAX}} + I_{\text{MIN}}} \quad (24.4)$$

By increasing both I_{MAX} and I_{MIN} , reduced the contrast occurs. This means that if I_{MAX} and I_{MIN} are increased by the same amount, effectively a bias offset, then the contrast is reduced. Consequently, even in the case where optical scatter within the optical system takes a uniform profile across the detector plane, a reduction in imaging performance occurs.

An example of veiling glare being encountered in an optical design was during the fabrication of a reasonably wide field of view prototype imaging lens which used a plastic aspheric lens on the front surface. To generate the proof of concept design at reasonable cost it was necessary to fabricate the aspheric profile using diamond machining, thereby avoiding the cost of the mold in the near term. It was found that contrast in the image was noticeably lower than expected. The problem was found to be due to veiling glare from the diamond machined surface. The good news is that when the lens was transitioned to a molded part, the problem was completely eliminated because the molded part surface roughness was far lower than that of the diamond turned part!

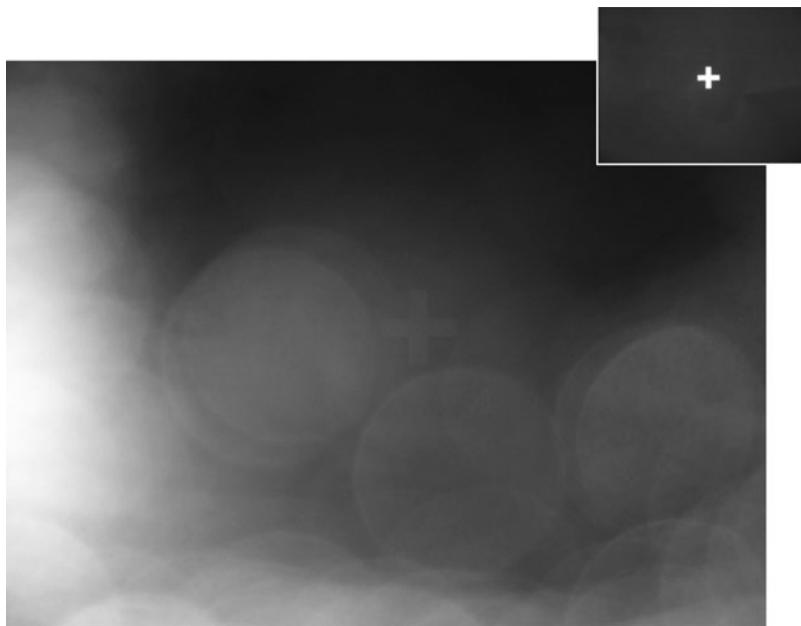
Cleanliness

Veiling glare can be caused by smearing contaminants and dust on front windows or near the pupil planes of a system. An example of this within a commercial grade digital camera can be seen in Fig. 24.9. In an optical system it is therefore wise to have a protective front window which can cope with regular cleaning without becoming damaged. For example, good quality single lens reflex (SLR) camera lenses often allow a removable UV filter to be mounted in front of the lens. This not only provides UV protection but also it serves as a protective and replaceable window. In rough environments use of a replaceable window component will prolong the overall system lifetime.

For any surface close to either the focal plane or an intermediate focal plane, special care must be taken to ensure good cleanliness. This is primarily to prevent dust particles from being directly imaged into the focal plane. In slow optical systems with small apertures cleanliness is a

Figure 24.9

Commercial Digital Camera Exhibiting Veiling Glare Due to Dust on Objective Lens. Inset Is the Image of the Diffusely Lit Test Scene—Black Felt with a White Cross (Obtain Focus)



priority. In such systems small dust particles may be a significant portion of the diameter of the pupil. The depth of field is large in a slow system, so several optical surfaces may contribute particle obscurations to any given field point. This may cause obscurations to occur in the image. The dust particles in slow systems are often likened to being “rocks in the image.”

Suppression Techniques

Three Bounce Rule

The key to scatter suppression is to create an opto-mechanical housing structure which maximizes the number of light absorbing surfaces that scattered light must strike before reaching the focal plane. If blackened housing surfaces, baffles, and thread pitches are correctly chosen, it is possible to effectively and with low cost, eliminate most sources of structural scatter. If there's one rule to remember, it's the “three-bounce rule.”

Three bounces of a 1% reflection will reduce the amount of light to *one millionth* of the original energy. This is a great rule of thumb for most scatter suppression applications!

Threads and Baffles

Threads can be an inexpensive and highly effective way of achieving the three bounce rule. However, the pitch of the thread is important. A thread can never be perfectly sharp, and so the tops of threads can be considered to be flat topped, effectively reflecting light back into the optical path. We can consider that the thread may not be much sharper than 0.025 mm on a top without requiring special fabrication and handling procedures. If the thread pitch is large, then the flat top frequency will be reduced along the length of a bore. Decreasing the pitch will conversely bring the flat tops closer together giving the bore more direct reflection surface area.

Choosing the angle of the thread also has some influence on scatter suppression performance. If the thread is too shallow, then three bounces are unlikely to occur when the light interacts with the thread. Furthermore, it is preferable to utilize where possible standard thread machining practices to minimize special machinist threading operations. A study of baffle placement has been provided in the next section.

Baffles can improve system performance significantly. The direct path analysis method discussed in “Scatter Path Analysis” section describes how to determine key locations for baffles. This is useful to identify primary scatter pathways and minimize debug time later.

A typical form of baffle commonly used in an optical system with an internal focus is a field stop. The field stop location is a natural break point in the optical design to mop up unwanted light.

TRADE-OFF STUDY—BAFFLE DESIGN The placement and number of baffles in an optical bore will influence the scatter suppression offered by those baffles. Figure 24.10 gives some examples of baffle structures within an optical bore. The relationship between the pitch of the baffle and the height of the baffle within a fixed diameter of bore is also important. For example, it is clear that a short baffle height (h) will require a smaller baffle separation or pitch (p) to affect a reasonable amount of optical scatter suppression. The relationship of (h/p) is a good metric to note when placing baffles within the structure.

Figure 24.10
Example of Baffle Structures in a Bore

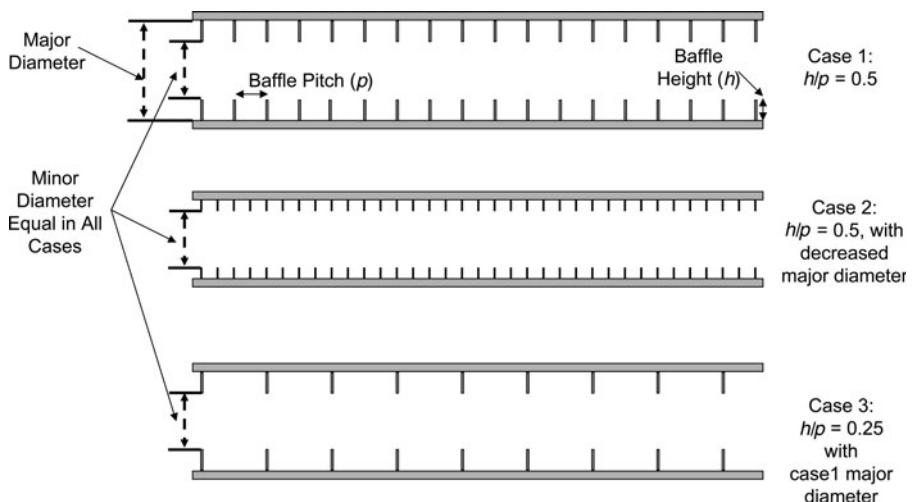
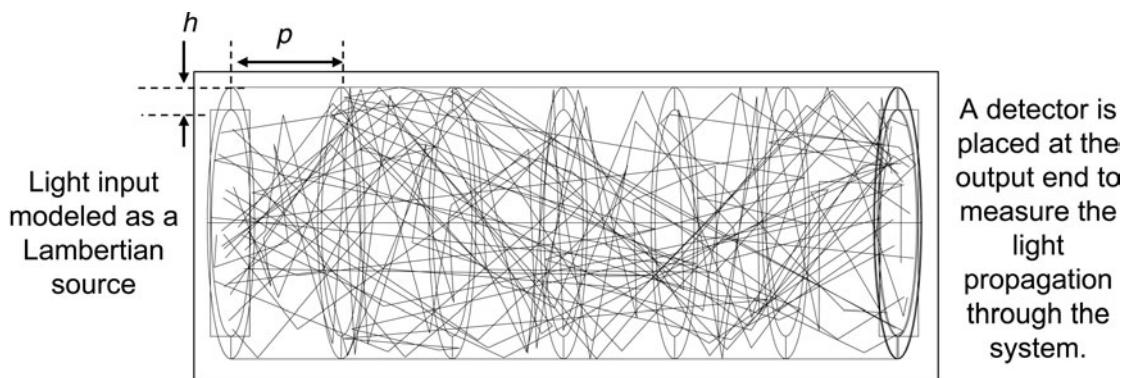


Figure 24.10 shows two examples where $h/p = 0.5$, that is, the baffle separation is double the baffle height. In the third example, a lower frequency of baffles provides an h/p ratio of 0.25. A lower value of h/p is likely to provide poorer scatter suppression.

To better understand the optimal spacing of baffles the following trade study was conducted:

A bore with no elements, containing only baffles was modeled. At one end was placed a Lambertian source and at the other, a single element detector. Baffles were placed in the tube with a fixed inner diameter and the spacing between them was varied; a baffle was always retained at the input and output face. The baffle major and minor diameters were fixed. Baffles were assumed to be infinitely thin, so top surface (flat-top) reflections are ignored. Diffuse and specular surface baffles profiles were both examined. The effectiveness of the baffles was assessed by examining the amount of scattered light which reached the detector. Results were normalized to the case of a perfectly specular baffle being located at each end of the tube. Light which passed directly to the detector was removed because it was not scattered. Light paths which experienced more than 100 bounces were eliminated. An example of the baffle test configuration is provided in Fig. 24.11.

The analysis shows that as more baffles are added and the pitch (p) between them reduces the amount of scattered light reaching the detector is reduced. The height of the baffle (h) is constant in the analysis.



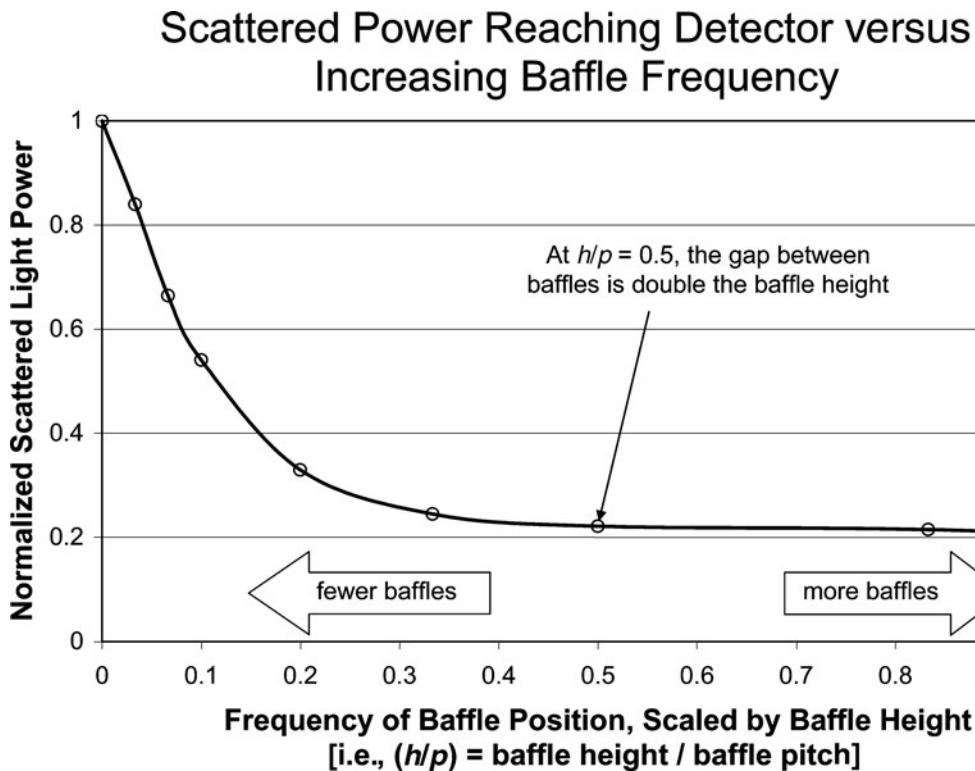
To be certain of the effectiveness of baffles in a real system, then a stray light model is required. In this example, a tube containing no lenses is modeled. The tube and baffles in this case are highly specular and reflective. This allows the placement of the baffles to be better considered in isolation of simple absorption. A highly diffuse version of the model is also considered.

Figure 24.11
Baffle Analysis Configuration

Results are expressed in terms of the amount of scattered light power which reaches the detector, shown on the y-axis, versus the ratio of the height of the baffle to the pitch between them (h/p), shown on the x-axis. Results are provided in Fig. 24.12 for the case of a specular baffle.

Specular baffles would not be recommended for use in a real optical system. Specular baffles were chosen to examine how the positioning and number of baffles can influence scatter suppression independently of the baffle material type. It is of little surprise that the diffuse baffles achieve better absolute energy suppression than the specular baffles. A brief study comparing specular and diffuse baffles is provided in Fig. 24.13. The diffuse material scatters 90% of the radiation into a 90° cone, and 10% of energy is absorbed at each bounce. The specular scatter object reflects 99% of energy specularly, losing only 1% through absorption at each bounce.

The rate at which scatter suppression occurs with reduced baffle separation can be determined by examining the slope of Fig. 24.13. Reviewing Fig. 24.13 indicates a similar trend for both specular and diffuse baffles. When the height to spacing pitch ratio (h/p) reaches approximately 0.5 little further scatter suppression achieved by the addition of more baffles.

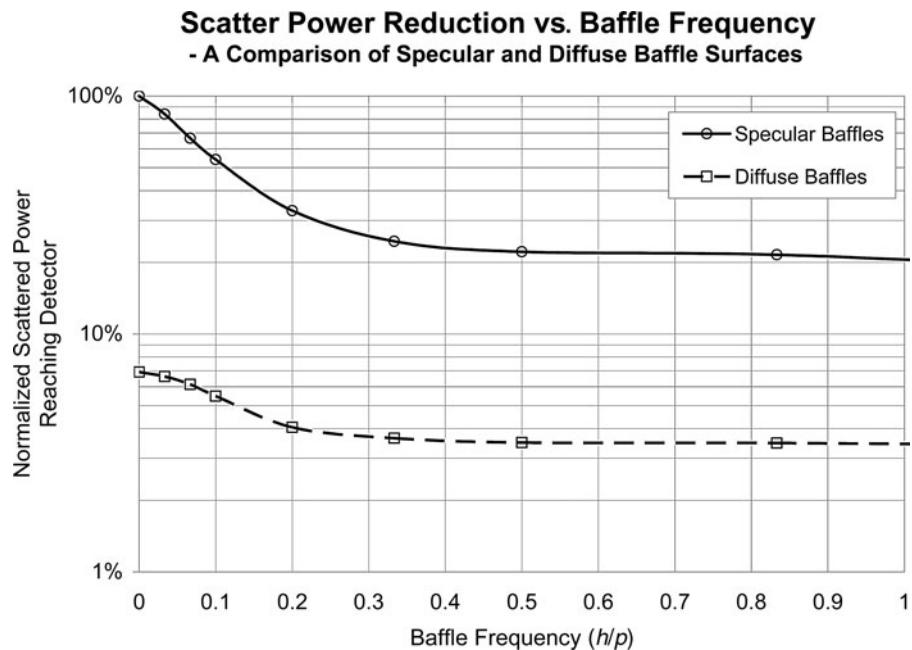
**Figure 24.12**

Analysis of Scatter Suppression with Frequency of Baffles along a Tube

The small amount of improvement offered by more baffles requires increasingly more complexity in the mechanical design. Furthermore, beyond $h/p = 0.5$, we move toward a scenario where the number of flat tops of the baffles would become significant; this treatment only considered infinitely thin baffles. The inevitable flat tops may begin to decrease baffle effectiveness. This is discussed further in “Thread Profiles” section for the case of threaded bores.

As discussed, this analysis assumes imperfect absorption from the baffles. Baffles are, however, typically good absorbers, or at least should be designed to be. If we make the assumption that the baffles are 100% absorptive, then an optimal baffle geometry using a minimum number of baffles can be defined. This is now considered.

Figure 24.13
A Comparison of Diffuse and Specular Baffles

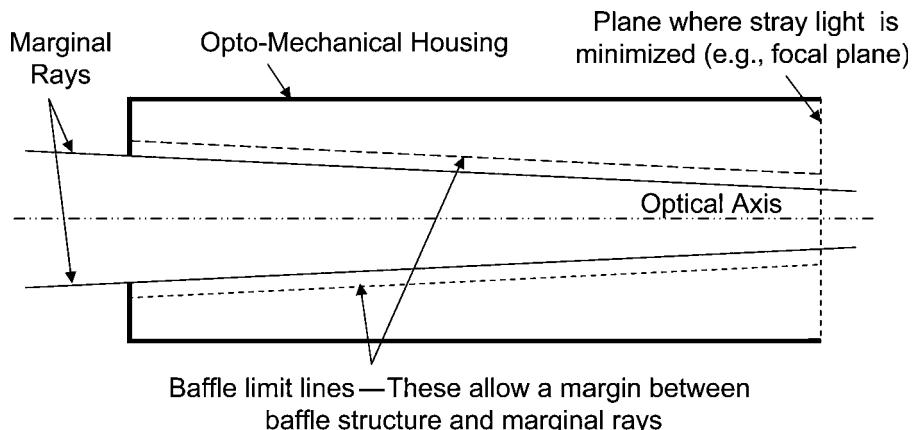


OPTIMAL BAFFLE GEOMETRY Assuming that the baffle surface can be made highly absorptive by surface roughening and blackening, it is possible to define an optimal baffle spacing arrangement. This uses the minimum number of baffles needed to block single structure reflections.

In Fig. 24.14 an optical system bore is shown, with the optical system marginal ray paths indicated. For simplicity, this is shown in a volume where there are no optical components. The optomechanical housing walls, the entrance aperture, and the marginal rays are defined. The focal plane is defined, but this could simply be an intermediate location where we want to ensure we have minimal stray light entering the optical system, that is, it could be the primary mirror location in a Cassegrain telescope system.

In Fig. 24.14 an allowance is provided for mechanical tolerances affecting the marginal ray locations with respect to the baffle locations. The baffles are, therefore, insets from the marginal ray line. The inset amount required will vary depending on the optical tolerances of the system. It is better to slightly undersize this baffle diameter at this stage. Erring on the side of caution is preferable because if the baffle diameter is too large the marginal rays will be vignetted.

Figure 24.14
An Optical System
Bore



To define the locations of the baffle, we work from the detector plane back toward the entrance aperture. The full focal plane is used in this description. The detector may, however, only cover the central portion of the focal plane. Use of the full diameter of the focal plane allows the baffling structure definition to be the most effective. Use of a final baffle immediately before the focal plane is common practice, not only for scatter suppression, but often such a component is used to assist in locating the detector with respect to the optical axis.

The direct line of sight paths to the detector plane are systematically eliminated from back to front. To define the location of the first baffle, “locator line 1” (LL1) is drawn, see Fig. 24.15. LL1 is the pathway to the housing wall. Note, that locator lines defined here do not necessarily describe a specular reflection off the housing wall (that is, reflected

Figure 24.15
Locator Line 1
Defines Ideal Position
of First Baffle

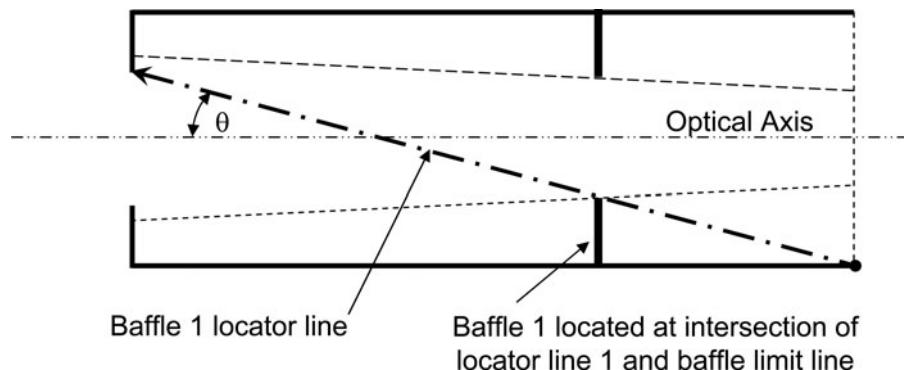
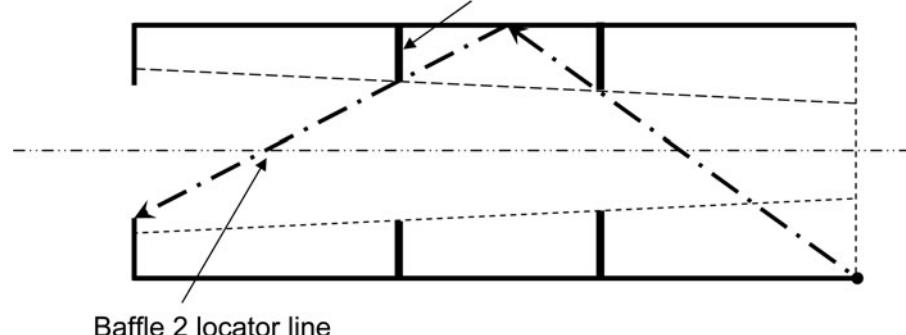


Figure 24.16
Locator Line 2
Defines Ideal Position
of Second Baffle

Baffle 2 located at intersection of locator line 2 and baffle limit line



angle does not necessarily equal incident angle in the case of scattered light). The direct path from focal plane to entrance aperture may be a component of a Lambertian scatter, for example.

Baffle 1 as defined on Fig. 24.15 limits the field angles which can directly reach the focal plane to a maximum of $\pm\theta$ about the optical axis. The sensitivity of the tolerances of the baffle location along the optical axis are determined by the steepness of θ , that is, $dz = dh/\tan(\theta)$ and this should be considered when insetting the baffle limit lines from the marginal rays.

The second baffle location is defined by the intersection of locator line 2 and the baffle limit line on Fig. 24.16. Similarly the third baffle is defined by the intersection of the locator line 3 and the baffle limit line on Fig. 24.17. Figure 24.18 demonstrates that a fourth baffle is not required in this case.

Figure 24.17
Locator Line 3
Defines Ideal Position
of Second Baffle

Baffle 3 located at intersection of locator line 3 and baffle limit line

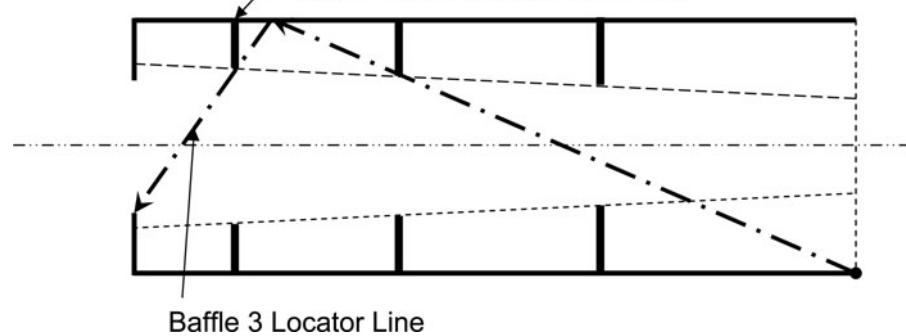
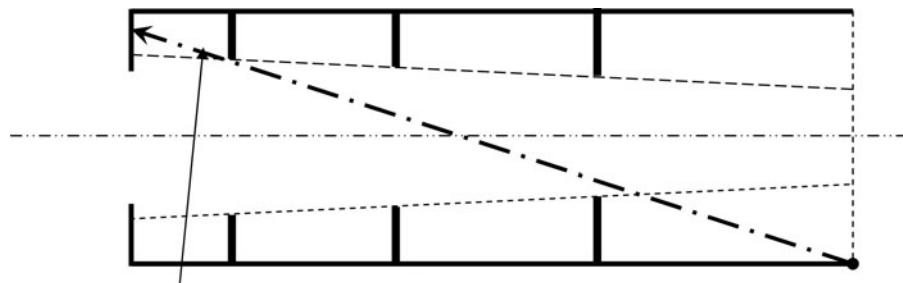


Figure 24.18

Locator Line 4 Indicates That No Further Baffles Are Required



Baffle 4 locator line indicates that further baffling is not required

If the marginal ray bundle within the housing is diverging, then more baffles may be required. This can be seen in Fig. 24.19.

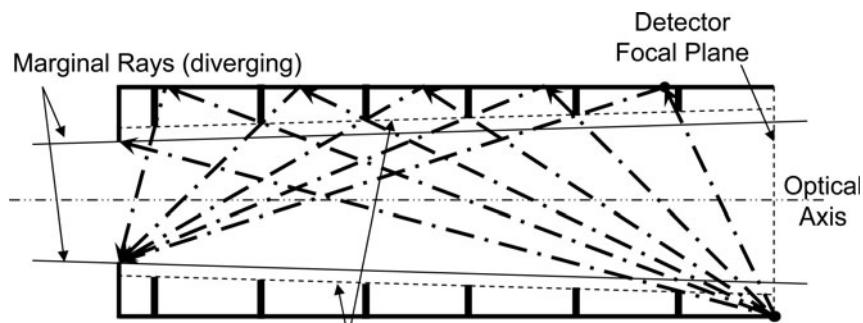
Lastly, light entering this optimized baffle arrangement which experience specular reflections, will retro-reflect back toward the entrance aperture. This uses up two surface bounces. In order to be again directed toward the focal plane, a third and possible fourth bounce would be required. This baffle optimization can therefore be considered as a technique to optimize the design to meet the "Three Bounce Rule" discussed earlier.

BAFFLE PROFILE DESIGN Yes—it is even possible to optimize the profile of an individual baffle! Figure 24.20 shows a typical baffle profile.

Baffles are thin plates which are spaced along the inside of the optical housing providing a clear volume defined by the minor diameter. The diameter of the tube wall is referred to as the major diameter.

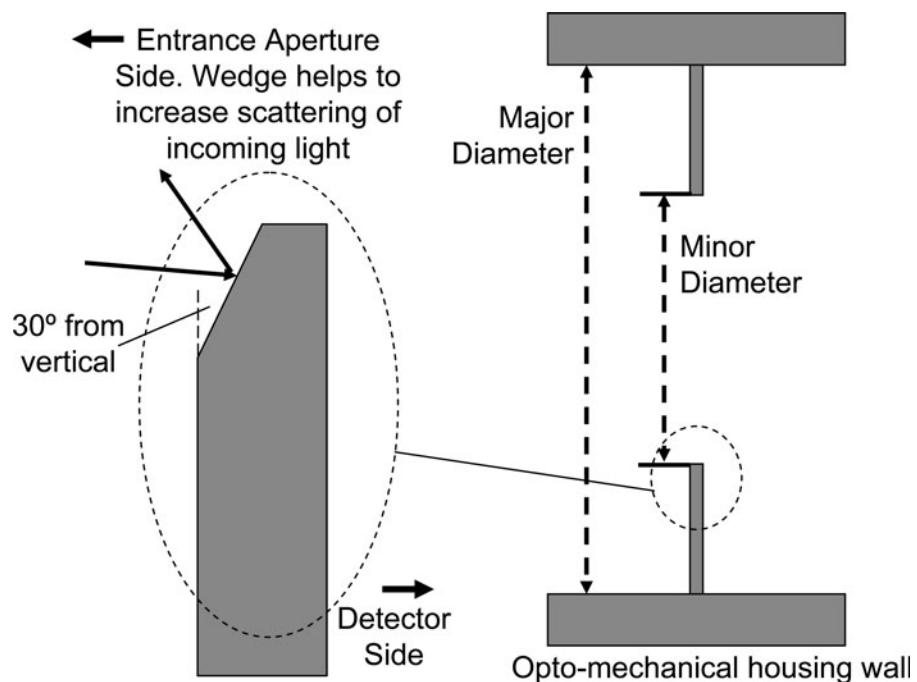
Figure 24.19

Diverging Marginal Rays Require an Increased Number of Optical Baffles



Baffle limit lines—These allow a margin between baffle structure and marginal rays

Figure 24.20
Typical Baffle Profile



Baffles have some limitations:

- They cannot be infinitely thin, nor can the tips be infinitely sharp.
- A reasonable machining limit for the sharpness of the baffle is in the range of 0.1 mm (0.004 in) to 0.25 mm (0.010 in).
- Sharper baffles become hard to make and are more difficult to handle.

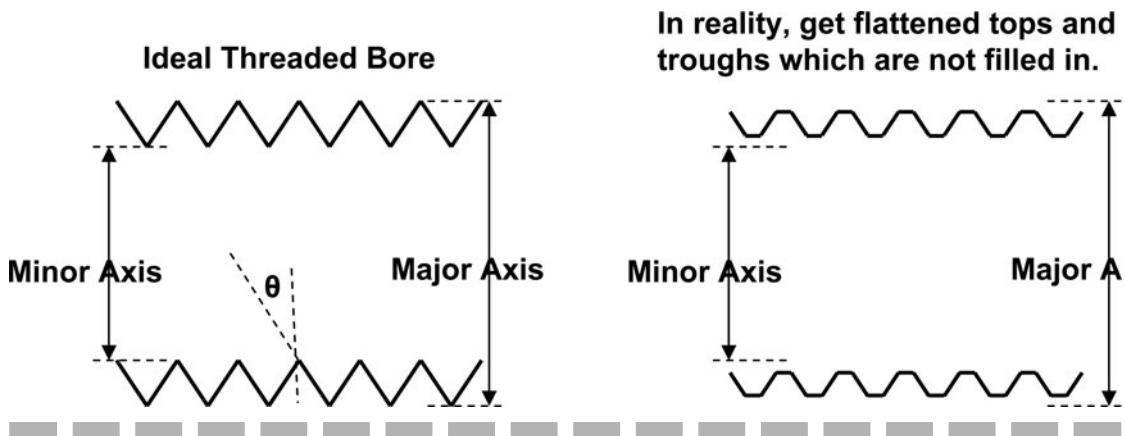
To avoid these difficulties it is often easier to have a baffle which is thicker, but reaches a point at the inner diameter. This can be achieved using a chamfer. The chamfered side of the baffle faces the entrance aperture. This helps to maximize the amount of light deflected into baffles or threaded structure in the walls. The effectiveness of this design profile would need to be modeled (or empirically evaluated) for any given case, but this is a good design practice applicable in most cases.

THREAD PROFILES Applying a thread to the inner bore of an opto-mechanical housing is an inexpensive technique for obtaining baffle suppression. Threads commonly applied to optical housings are in the range of 24 to 32 threads per inch (tpi), or 1 to 1.25 threads per millimeter (tpmm).

Typical machine threads have a pitch of between 45° to 60° , although custom threads can be generated. A typical thread profile is provided in Fig. 24.21a. In practice threads cannot be perfectly sharp. Figure 24.21b illustrates a typical thread, which will have a flattened top and a valley.

It is reasonable to expect the width of the flat to be in the range of ~ 0.1 mm (0.004 in) up to 0.25 mm (0.010 in). Care needs to be taken not to increase the pitch (that is, increase the tpi) of the thread too much. This will increase the total surface area of the flats, and hence lead to a surface which is overall more specular in nature.

A 45° thread angle provides a height to spacing (h/p) ratio 0.5, which agrees well with the baffle trade-off study optimal separation discussed in "Trade-Off Study—Baffle Design" section, where an h/p ratio of 0.5 was found to be optimal. Retaining the same h/p ratio of 0.5 using a 45° slope, but with a larger pitch increases the depth of the thread. Using an increased thread pitch will therefore require a large bore diameter. For this reason, threads of 24 to 32 tpi are preferred in most optical housings where space and volume are at a premium. A typical call-out for a thread on a drawing would be: "Cut antiglare thread 32 tpi" (1.26 tpmm). This note tells the machinist that this thread doesn't mate to another part. Note that at 32 tpi (1.26 tpmm) with a flat tops and valleys of say 0.002 in (~ 50 μm), the ratio of flat spots to slopes is $1:1/(2 \times 0.05 \times 1.26) \approx 1:1/(2 \times 0.002 \times 32) \approx 1:7.8$, that is, approximately 12.8% of the bore will be acting as a near flat surface. The peak portion of this (that is, excluding the valleys) is half of this (6.4% in this example) and will be visible when



Figures 24.21
Ideal and Real Thread Profiles

directly viewing down the barrel. The valley portion increases the likelihood of scatter bouncing out of the thread furrow before three bounces have been achieved. If the peaks are not highly absorptive then stray light will be propagated on through the system from these peak sites.

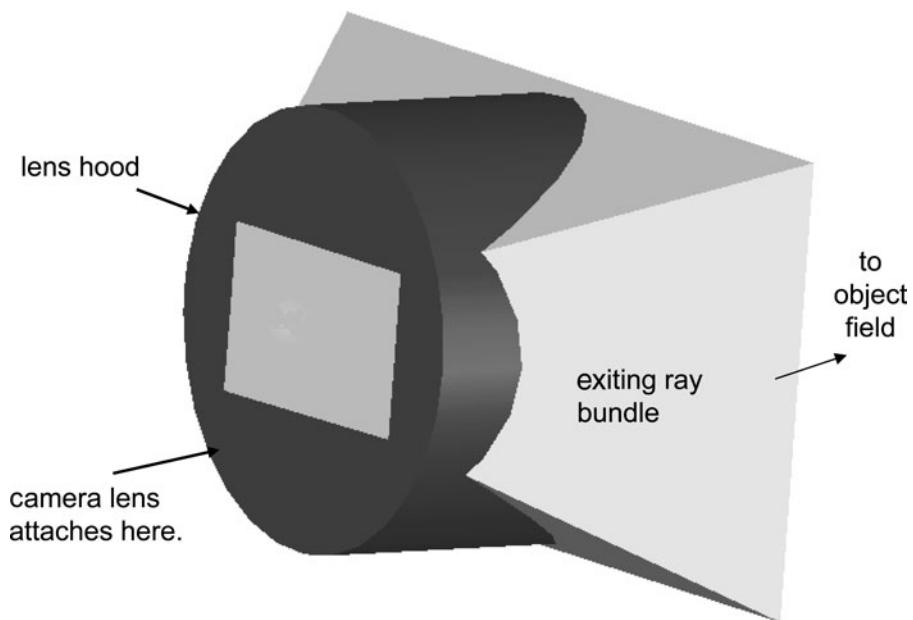
Tighter thread spacing will make this ratio of peaks and valleys to total bore length worse as threads will get closer together, but the flats and valleys will not become smaller.

Threads are usually used to augment scatter suppression in an optical system but usually some form of baffling is required to optimize scatter suppression. Threads can be considered helpful, but not a cure all, for optical scatter suppression.

Lens Hood Design

A lens hood is a very effective way to remove unwanted veiling glare from an optical system. An example of a fixed hood for a reasonably wide field of view lens is given in Fig. 24.22. As is typical of wide angle

Figure 24.22
Lens Hood Design



Pupil separation at front aperture of lens is large. A near rectangular ray bundle profile is described. This has been propagated to define the required hood characteristics.

lenses, the field bundles are largely separated at the front surface. The marginal rays of the system define a near rectangular shape at the front lens, this is in effect a projection to the rectangular aspect ratio of the sensor. The rectangular projection expands in size from the first lens surface toward the object. A conic section has been intersected with the image field marginal rays to determine the optimal hood shape. This maximizes veiling glare protection whilst ensuring that the fields required are not vignetted.

Care has to be taken to sufficiently oversize the hood profile to cope with manufacturing tolerance variations associated with the lens fabrication and assembly and of the hood fabrication and positioning onto the lens. On zooming systems the effectiveness of the hood is limited by the wide angle field of view. If the same hood is used in the narrow and wide field of view, then the narrow field of view will be sub-optimally protected. The narrower field of view can have a longer hood before vignetting of the edge of the field rays occurs. This shortcoming may be minimized if the hood changes shape when zooming, or if the objective retracts into the housing when zooming. However, this requires significant opto-mechanical design effort. It is sensible to make the inner bore of the lens hood matt black, or threaded black if possible.

If the optical engineer requires that a hood be used in the optical design, then it really does need to stay. Don't let the marketeers cut it back because they think it looks better without it (see Chap. 25)!

Material Characteristics

Material scatter characteristics can be well defined by *Bi-directional Reflectance Distribution Function (BRDF)* profiles. The *Transmission* version BTDF is typically used to characterize optical materials. An example of a BTDF profile is provided in Figs. 24.23 and 24.24.

Figure 24.24 provides the scatter characteristics of a glass aspheric surface fabricated on N-BK7 glass. The other side of the element is a spherical lens which contributes little to the overall scatter profile in this case. The measurement is made by scanning the profile through a range of angles using detector mounted on a goniometer. Figure 24.25 describes a goniometer with a one-dimensional scan-axis apparatus. In the measurement example provided the source was a laser beam at

Figure 24.23
BTDF Profile for an
Aspheric Surface
Fabricated on N-BK7
Glass

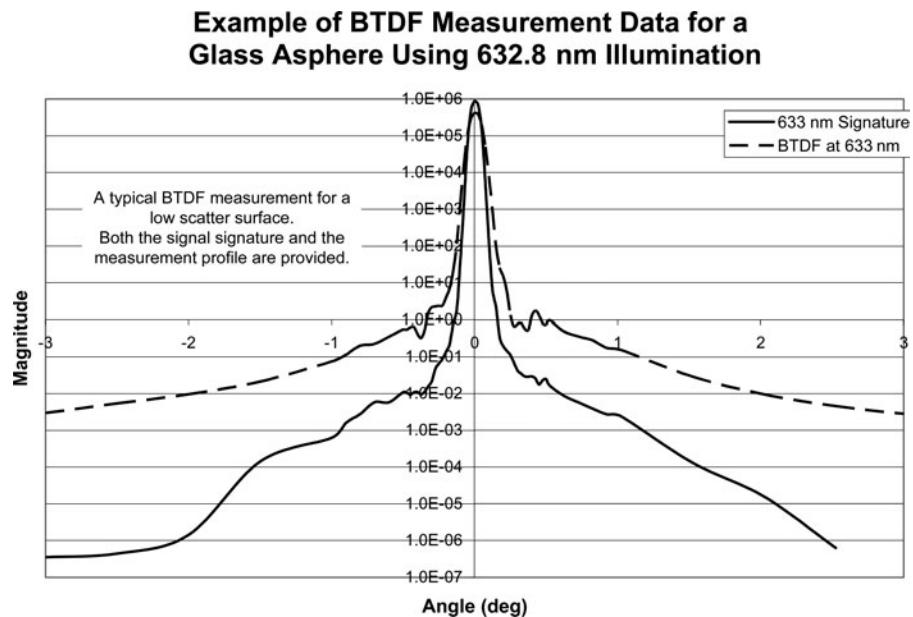
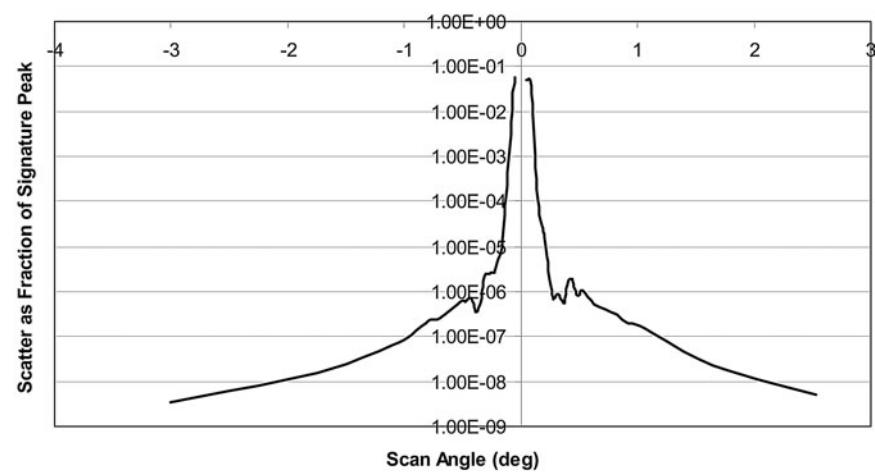


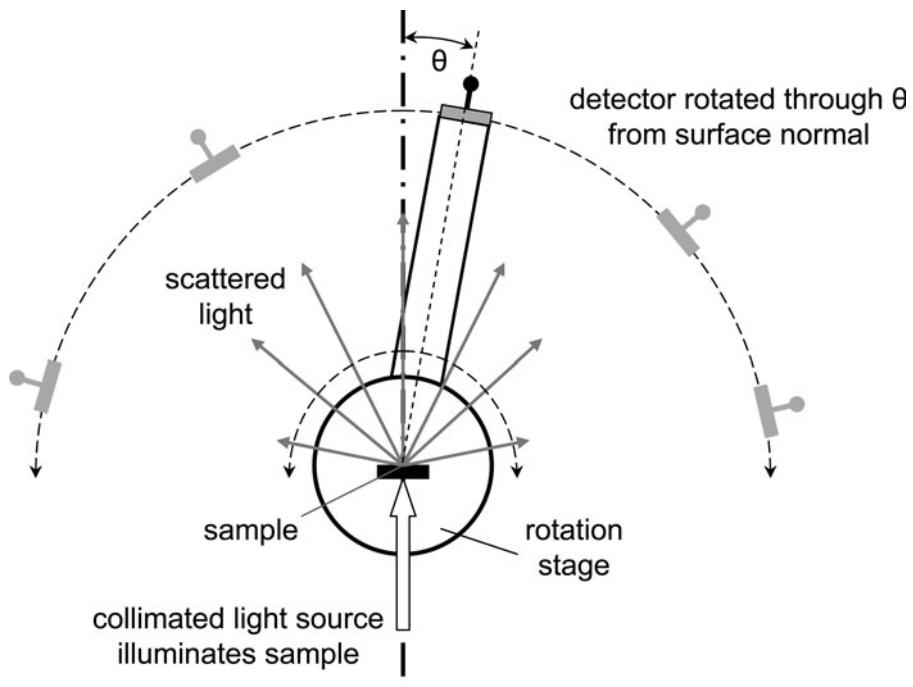
Figure 24.24
BTDF Profile with the
Signal Component
Removed

**BTDF for Glass Asphere Surface at 632.8 nm.
Scatter as Fraction of Signature Peak**



BTDF with signature peak removed. Data was not recoverable on axis because computational error became too large.

Figure 24.25
One-Dimensional
Goniometer Scatter
Measurement System



632.8 nm. This example shows a relatively narrow scatter angular profile and so the range of angles sampled was small. The large central peak has to be removed, and the revised profile is provided in Fig. 24.24.

A coarse surface profile will generate optical scatter over a full hemisphere (for an opaque) material. A transmissive scattering object will generate scatter for forward and backward, over full sphere.

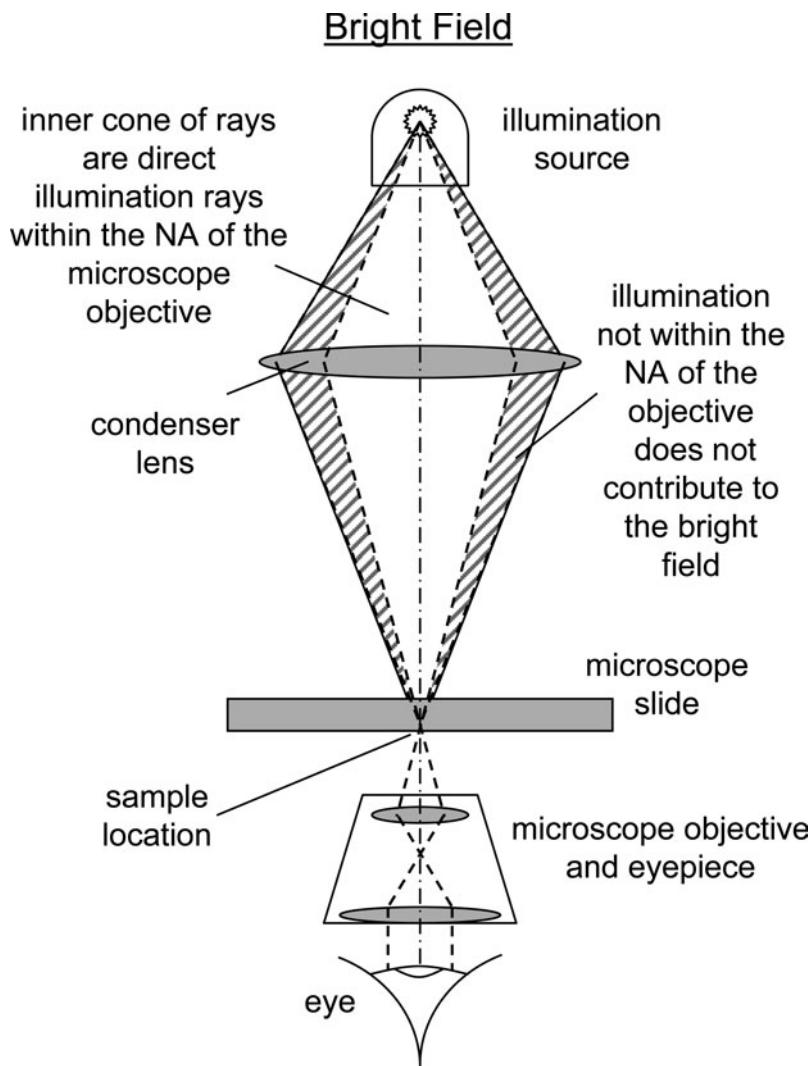
It is noted that in the surface scatter profile can be integrated. Examination of the TIS can be used to determine the surface roughness.

Bright Field and Dark Field

Optical scattering is not always encountered as an unwanted side effect. It is possible to use optical scattering beneficially to obtain improved contrast imagery.

Consider an example scenario, where an optical microscope is being used to image a biological contaminant found on an optical surface. Figure 24.26 describes a configuration where the sample is illuminated directly. The illumination source fills the numerical aperture (NA) of the microscope. This generates bright-field image of

Figure 24.26
Bright Field
Schematic



the sample. An example of an image obtained in the bright field is provided in Fig. 24.28. Bright field imagery is suitable when viewing sample where the transmission through the sample varies. The variation in the transmission profile generates the contrast in the observed image.

Figure 24.27
Dark Field Schematic

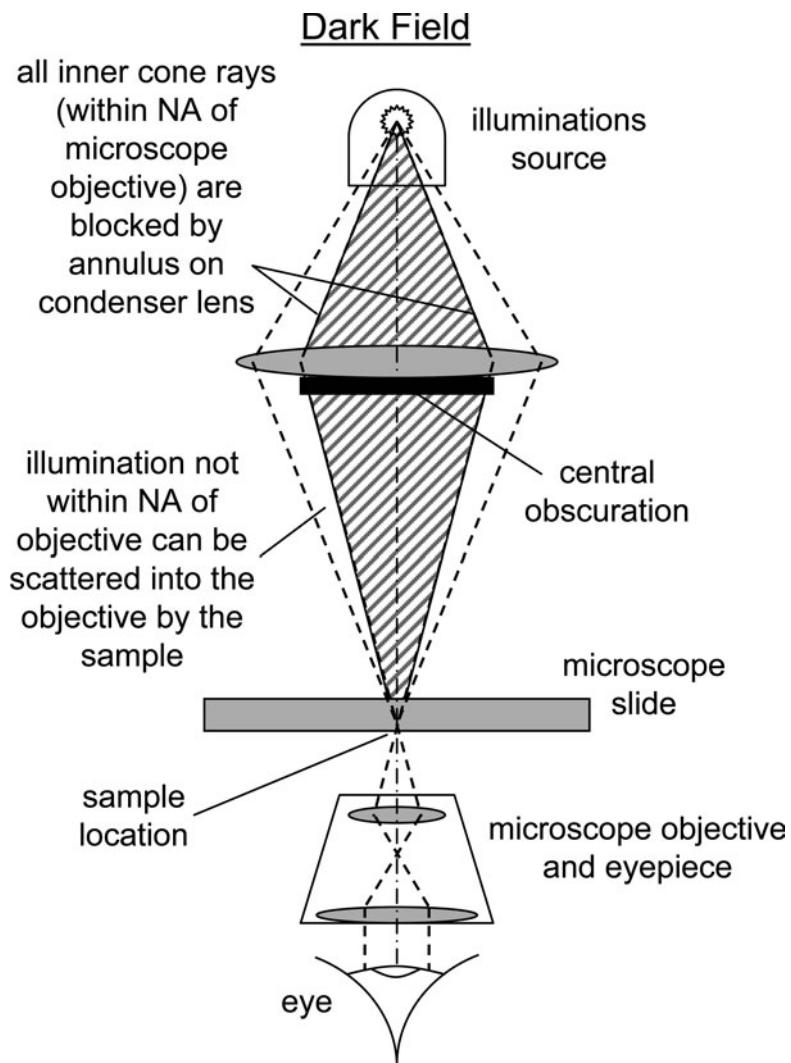


Figure 24.26 shows exactly the same sample, viewed with the same microscope, but the illumination cone has been altered. Here the illumination profile is an annulus, all of which lies outside of the NA of the microscope objective. Light will only enter the microscope if it is scattered back into the NA cone of the microscope objective. If the sample is removed, then the field will be completely dark, hence this is known as dark field imaging, or dark field microscopy. An example of an image obtained in the bright field is provided in Fig. 24.29. This type of illumination is useful for observing objects which exhibit low optical transmission variation making bright field imagery less suitable.

Figure 24.28

Bright Field Image Example

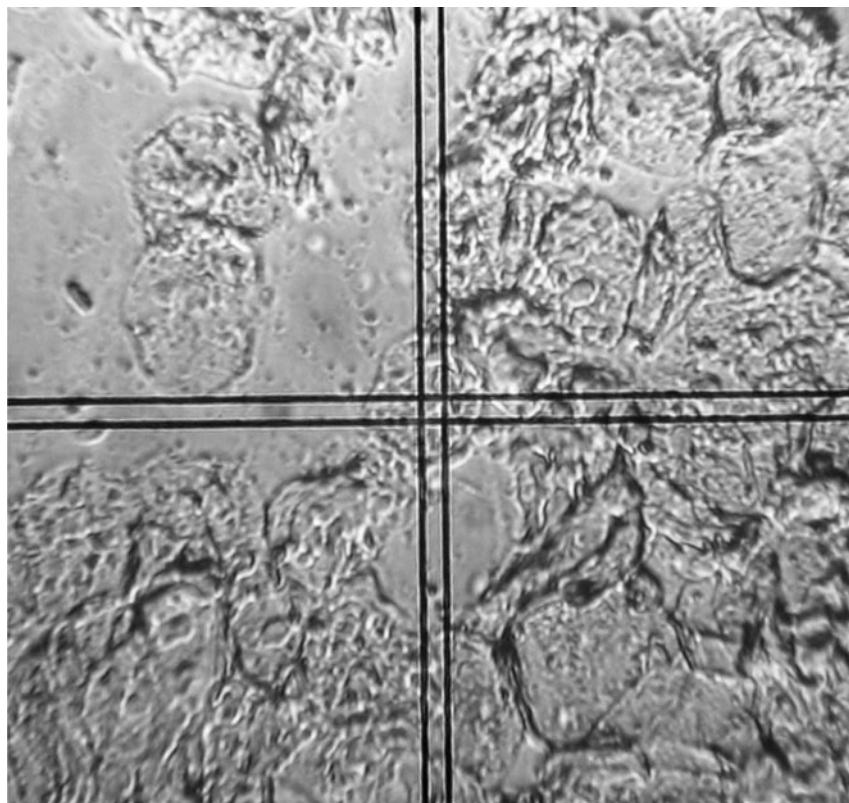
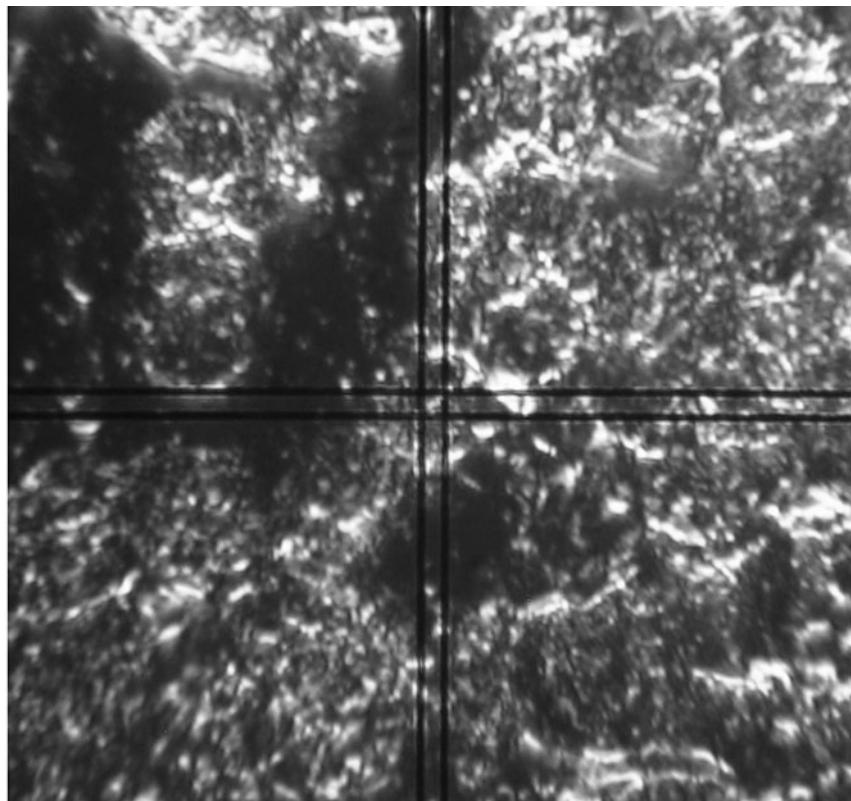


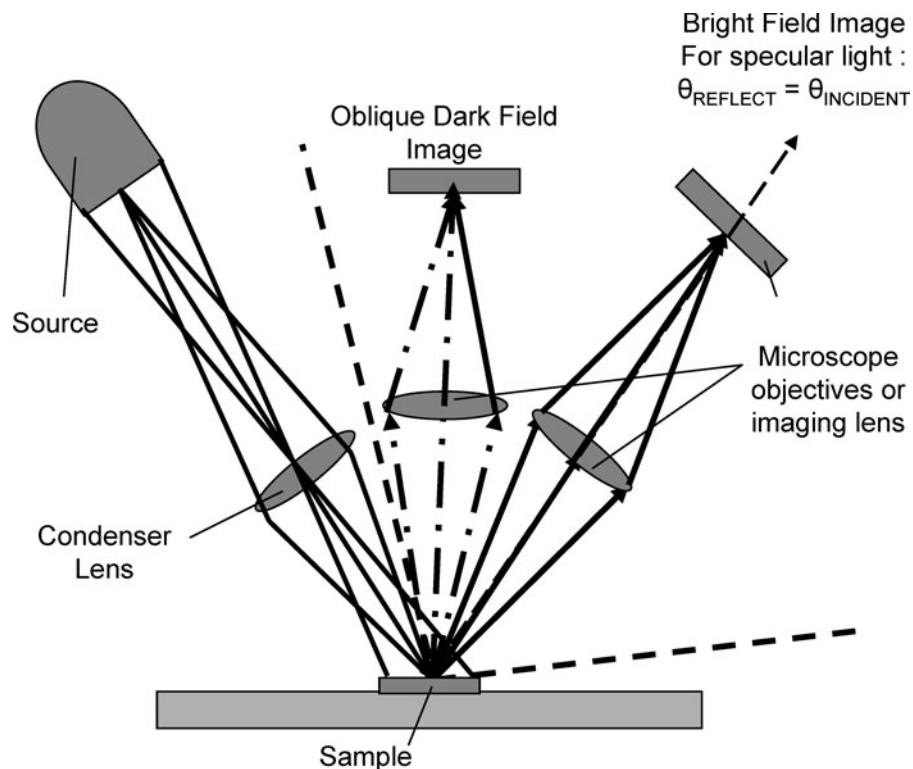
Figure 24.29
Dark Field Image
Example



Obtaining both bright field and dark field imagery of a sample will often increase the overall understanding of the sample being viewed.

Figure 24.30 shows both bright field and dark field imagery. The bright field is achieved using the specularly reflected portion of the illumination cone. The dark field imagery is obtained using the diffusely scattered portion of the illumination cone. The dark field image uses an oblique illumination profile which is inherently asymmetrical. This is unlike the example provided in Fig. 24.27 where the illumination profile is a symmetrical annulus.

Figure 24.30
Bright and Oblique
Dark Field Imaging
System



How to Avoid Unwanted Stray Light

The following list summarizes tried and tested techniques to avoid scatter problems in an optical system:

- Use the three bounce rule!
- Check single bounces in the optical design.
 - Conduct single bounce analyses in IR systems to identify Narcissus contributions.
- Do not oversize optical elements within the optical system by more than is necessary to account for optical tolerances and mechanical mounting.
- Employ threads, baffles, and field stops where possible.

- Identify the key locations for thread and baffling positioning.
 - Use direct modeling to identify key stray light paths.
 - Ensure time for a prototype evaluation of stray light performance in prototype first article units. Empirical testing is the ultimate proof of success!
- Be careful about the thread pitch and/or baffle spacing. A pitch (or separation) to height ratio of 2:1 is optimal in most cases.
- Design the baffle locations to minimize stray light from direct ray paths.
- Blacken as many surfaces as possible.
 - Metallic surfaces should be black anodized as a minimum.
 - Consider using shedding free flocking materials. These can be applied for low costs and be highly effective. Consideration of the adherence of such materials may depend on the environment used.
 - Blacken lens edges and out of clear aperture surface mounting flats.
- Use AR coatings to minimize Fresnel reflections, especially in IR systems where higher refractive index materials are common.

Bibliography

- Fischer, R. E., Tadic-Galeb, B., *Optical Systems Design*, McGraw Hill, 2000.
- Hecht, E., *Optics*, 2nd Edition, Addison Wesley, 1987.
- Smith, W. J., *Modern Optical Engineering: The Design of Optical Systems*, 2nd Edition, McGraw-Hill, NY, 1990.
- Wyatt, C. L., *Radiometric System Design*, MacMillan Publishing Company, 1987.

This page intentionally left blank

Bloopers and Blunders in Optics

This section is presented in the spirit that we all learn from our mistakes and/or the mistakes of others. The truth is that none of us is perfect, and from time to time even the best of us make mistakes. If we can share, in the right spirit, these mistakes, we will all learn, and our industry will improve. We are careful not to use names or affiliations in any of the following material.

■ ■ ■ Distortion in a 1:1 Imaging Lens

A lens that is fully symmetrical on both sides of a central aperture stop will be free of all orders of distortion, coma, and lateral color. This is because precisely equal and opposite amounts of these aberrations are introduced on each side of the central aperture stop therefore producing a net zero aberration at the image. Some years ago, a lens was required which imaged from a convex curved CRT onto a flat ground-glass image surface. The lens needed to have less than 0.25% of distortion. The lens was designed to be completely symmetrical about its central aperture stop. Only after the lens was assembled and tested did it become apparent that there was a residual distortion of several percent. This was never

checked during the design effort because it was *assumed* that a fully symmetrical lens had zero distortion. The flaw in this assumption of symmetry was that the curved object surface immediately made the lens nonsymmetrical. This caused the distance from the edge of the field to the entrance pupil on the object side to differ from the corresponding distance from the edge of the field to the exit pupil, a clearly asymmetrical situation that will lead to distortion. If you can take advantage of symmetry, make sure that your system is fully and completely symmetrical!

We illustrate this in Figs. 25.1 to 25.3, where we show layouts and lens performance data for three different designs in which each of the lenses is fully symmetrical, except for the object radius which is shown as infinite, 100-mm convex, and 100-mm concave, respectively. The distortion is identically zero for the flat object design (the lens, object, and image are all completely symmetrical), and +1.4% for the 100-mm concave object and -1.2% for the 100-mm convex object.

Zoom Periscope

A new periscope was designed for the U.S. Navy many years ago. There were various optical innovations in the system, including anomalous dispersion glasses for improved color correction. One innovation was to

Figure 25.1
Fully Symmetrical 1:1
Magnification Lens

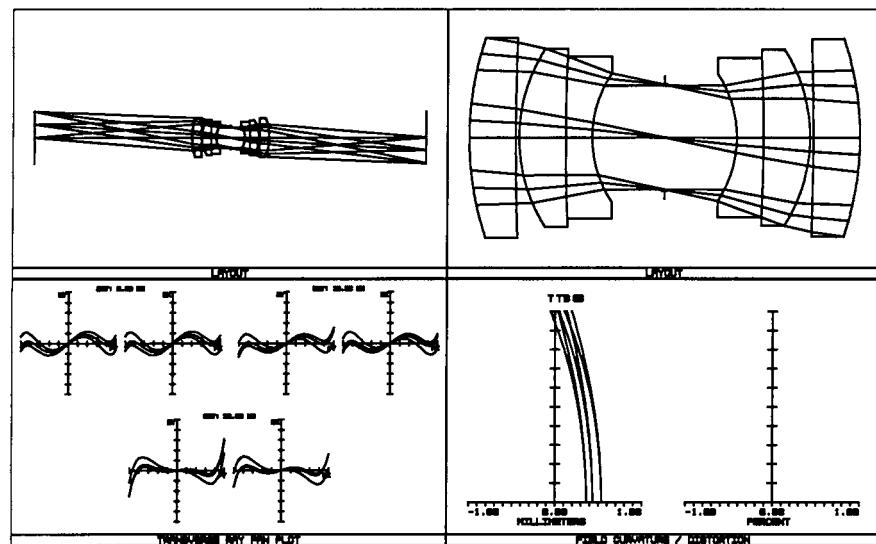
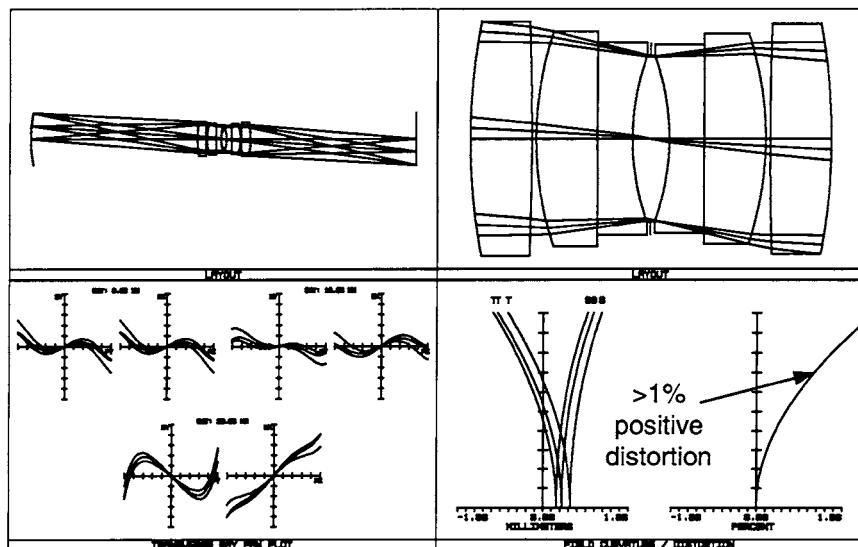
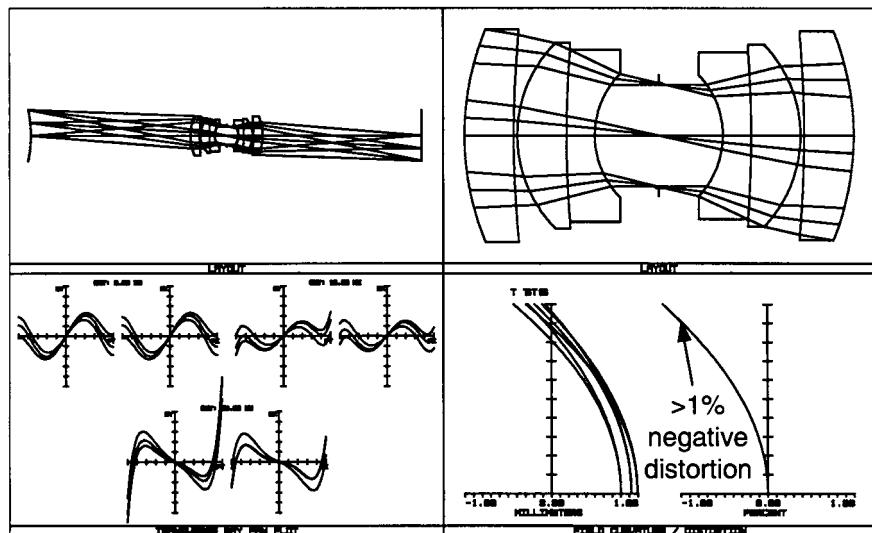


Figure 25.2
Symmetrical 1:1 Magnification Lens with 100-mm Concave Object and Flat Image



provide a continuous zoom lens to replace a $3\times$ discrete field-of-view change. While technically a success, during sea trials the person at the helm became disoriented during docking procedures while zooming the periscope, and the submarine crashed into the dock, causing major damage. The decision quickly was made to freeze the zoom and revert back to a discrete field-of-view switch. While this is not an optical problem as such,

Figure 25.3
Symmetrical 1:1 Magnification Lens with 100-mm Convex Object and Flat Image



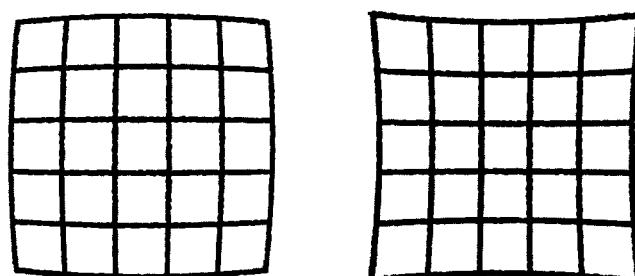
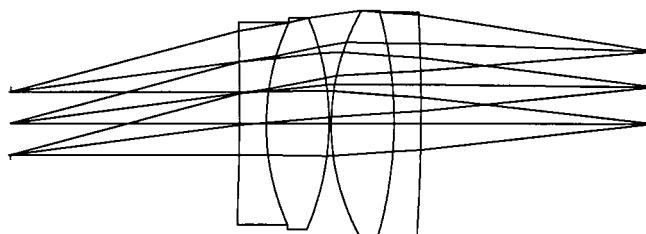
it is indeed a human factors and human engineering issue. Generally, the optical designer is quite remote from the human factors issues; however, if you ever come across a similar situation in your future work, be bold and bring it up. After all, if you don't, perhaps no one else will either!

Sign of Distortion

Generally, ray tracing an optical system from one direction or the other will yield virtually the same results with respect to image quality (note that in complex systems the results may not be precisely identical). There are, however, some significant effects relating to distortion which can take on a totally different form, depending on which way light is traveling. This is a difficult concept to grasp so we will illustrate it with a real design for an eyepiece.

Consider the Plössl form of eyepiece shown in Fig. 25.4. This eyepiece is designed to cover a full diagonal field of view of 40° , which is rather large for this design form. One result of this wide field of view is large distortion, at approximately 10%. If we ray trace from the eye to the image, we will predict negative or barrel distortion, as shown in Fig. 25.4. Now we will reverse

Figure 25.4
Plössl Eyepiece
Showing Sign
Reversal in Distortion
Depending on
Direction Lens Is Ray
Traced



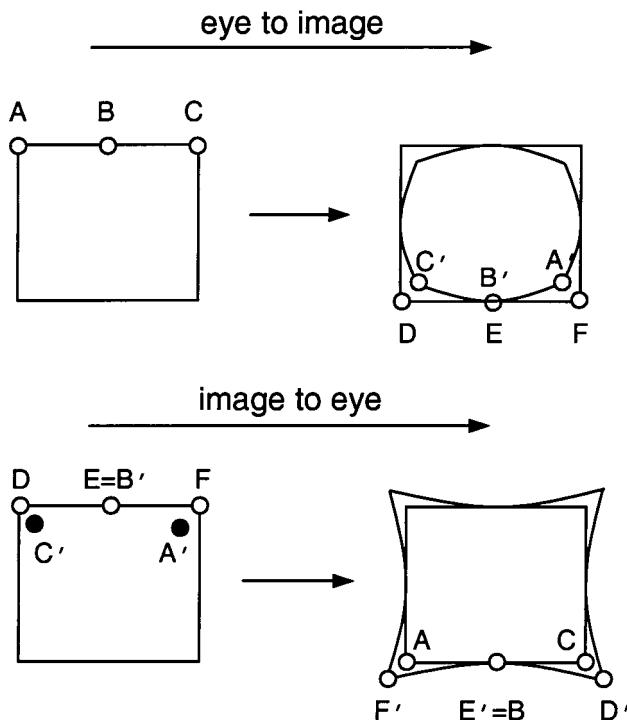
evaluated from
eye to image

evaluated from
image to eye

the design in the computer and ray trace from the new object plane, which used to be our image, into the eye. The resulting distortion will be similar in magnitude; only it will be reversed in sign! This is very interesting indeed and is difficult to understand. The following explanation should be sufficient: Think of distortion as being analogous to spherical aberration of the chief ray. After all, it really is similar to this. Now if we are ray tracing from the eye to the image being viewed such as an LCD display in an HMD application, then the lens elements will bend the chief rays more severely than paraxial optics will dictate, meaning that the off-axis chief rays will end up closer to the axis than their paraxial counterparts, resulting in negative or barrel distortion. Now let us reverse the eyepiece and ray trace from the image into the eye. The off-axis rays will bend more severely than their paraxial counterparts just like before; however, this will result in a *greater* ray angle entering the eye, and the resulting image will thus appear to the user as having positive or pincushion distortion.

Another way to explain and understand the opposite sign of distortion when tracing the rays backward is the following: When we trace a rectangular object (ABC) from the eye as in Fig. 25.5, it will be imaged into $A'B'C'$,

Figure 25.5
Explanation of
Distortion Sign
Reversal Illusion



where A' and C' have smaller heights than they should have if there were no distortion. Now, if we trace the rays backward, and if we take as the object the same points $A'B'C'$, we would end up with the angles A, B, C at the eye. However, when evaluating distortion, one always traces a rectangular object (or a regular-shaped object which is not distorted). In our case the object is DEF , where E is the same point as B' , and D and F are the points with larger height than A' and C' . Therefore, the angles after ray tracing to the eye corresponding to points D and F are D' and F' , and they are larger angles than A and C . E' is, of course, the same as B . Therefore, the key is that we do *not* take the same conjugate points in two ray traces.

The message here is to be extremely careful in assessing your performance, especially distortion. It would be quite disturbing if you predicted a given amount of negative distortion, only to find that the hardware produces the opposite sign.

Lens Elements That Are Not Necessary

An extremely weight-sensitive lens system was designed. The housing was manufactured of titanium and every gram needed to be accounted for. During the final design phase, one element became nearly flat with very long radii on each side. This element was about 10 mm thick. During the testplate fit, first the longest radius was made flat, and then the remaining radius of the element was made flat. This weight-sensitive lens system had, in effect, a flat window in its middle! In order not to look too foolish in front of the customer, one side was coated with a bandpass filter, which was originally planned to be coated onto one of the other elements. In the final design, this flat element was labeled “bandpass filter,” and everyone was happy. If elements are not serving a real function in a design, remove them.

Pupil Problems

Many years ago, a lens system was designed to reimaging high-resolution film onto a rear projection screen for viewing. As part of the system specifications it was necessary to provide an approximately $3\times$ zoom of

the central area of the film onto the screen. In order to “save money,” it was decided to use all off-the-shelf optics. The optics consisted of a zoom lens and at least one relay lens group. After months of mechanical design and system integration, the initial imagery proved to be excellent. Unfortunately, when the zoom was initiated, the image became darker and darker until it was totally black prior to reaching the high magnification. The problem was that as zoom lenses are zoomed, their entrance and exit pupils translate axially, sometimes by large amounts. What had happened was that, at the high magnification, the light from the exit pupil of the first lens module or group simply did not get through, or even into, the entrance pupil of the next lens module. It took many months and was quite costly to remedy the problem.

Not Enough Light

A machine vision system was designed some years ago to provide a $50\times$ magnification from the object to the image, with a CCD chip located at the image. The specifications called for a relatively long working distance and for a working *f*/number of *f*/3 at the object in order to be able to sense the *z* location of the surface under test (in the focus direction). First-order optics tells us that the final *f*/number at the CCD will be *f*/150, which is quite high. The customer was informed that there may be an illumination problem and that there may not be enough light. The reply was “no problem... we have been there before, and we can simply turn up the rheostat.” Eight months and several hundred thousand dollars later they did not have enough light and the project was canceled!

How can you prevent this from happening to you? We have two suggestions: First, carefully work through the radiometry and derive the required irradiance in watts per square centimeter on the CCD chip. These data, along with the data sheet for your CCD device, should allow you to compute your signal-to-noise ratio, which will give you a level of confidence in your having enough light. Another approach is to perform an empirical experiment. Set up your object and your illumination just as you plan to implement it in your system. Now take a normal CCD camera lens and use it to reimagine your object from some reasonable distance such as 0.5 to 1.0 m or thereabouts. The most important factor here is to now place a small circular aperture (a round hole in a piece of black paper is fine) in front of your lens in order to create the *f*/150

that you will have in your real system. If your camera lens has a focal length of 20 mm, for example, the aperture will need to be $20/150 = 0.133$ mm in diameter, a very small diameter! If this is difficult to obtain, you can attenuate the light to the required level by using neutral density filters along with, or instead of, a small aperture. It is important to emulate the ultimate irradiance on the sensor. If you then determine that you have enough light, you can proceed ahead with your system design. If not, you have work to do! Empirical tests, such as described here, are extremely valuable, and often they are simple to execute.

Athermalization Using Teflon

Athermalization can be a serious problem, especially, but not limited to, thermal infrared systems where the change in refractive index with temperature (dn/dt) is large, such as for germanium where the value is $0.000396/\text{ }^{\circ}\text{C}$. A system was built some years ago for the near infrared (IR) (just below 1- μm wavelength), and a bimetallic housing structure was utilized in order to maintain acceptable imagery as a function of temperature. Unfortunately, the required motion was larger than could be accomplished with typical housing materials, and Teflon was used as a spacer material in order to control one of the critical airspaces for athermalization. Initially, the system worked perfectly; however, it was later found that Teflon had hysteresis in its expansion characteristics, and when ambient temperature was restored, the system was out of focus. Ultimately, a more complex bimetallic housing using different metals solved the problem. If polymer materials are used for athermalization, do so with extreme care and do not ignore the hysteresis factor.

Athermalization Specifications

In a thermal infrared MWIR system, the airspace between a zinc sulfide and a zinc selenide element was very accurately controlled using a bimetallic housing structure in order to maintain focus through a wide temperature range. Initial tests in a thermal chamber showed the focus to be perfectly maintained. Several weeks afterward, it was discovered that from the outset of the project, refocus was permitted, and athermalization was not at all required. Read your specifications carefully!

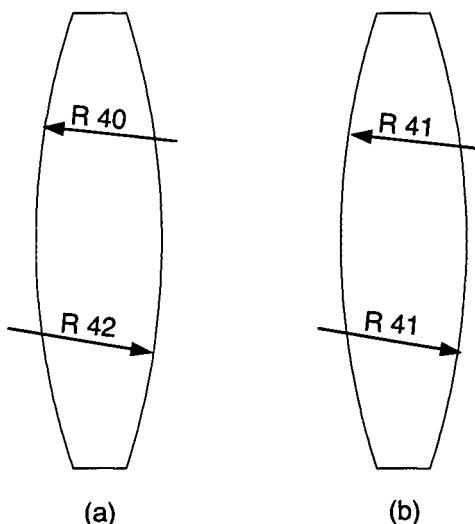
Bad Glass Choice

A very compact telephoto lens was designed for a production system. Because of the demanding packaging and optical performance, SF58 glass was used. The first problem was that the optical shop could only find $\frac{3}{4}$ ft³ of the glass in the world, not enough for full production. However, another problem arose after the first several hundred systems were completed—they all failed their MTF specification, especially on axis where astigmatism was present. After several weeks of intense study, it turned out that the SF58 elements had slumped during the coating operation. The lenses were supported in the coating chamber on rails, and the technician had been instructed to “set the temperature gauge to the red mark,” since that is the temperature where they coated all of their elements. Unfortunately, SF58 has the second lowest transformation temperature in the entire catalog, and at the temperature in the chamber, the elements softened just enough to slump a little, and that was enough to introduce the astigmatism. Fortunately, the elements could be fine ground and repolished prior to recoating at a lower temperature. A short postscript: the design was reoptimized using SF6, a much better glass, and the performance was virtually the same as with SF58.

Elements in Backward

This problem is far more common than it should be! At one of our short courses, 40% of course attendees' hands shot up in the air in response to the question “who has had the experience of elements being mounted backward?” Figure 25.6 shows a scale drawing of two 25-mm-diameter elements, one with radii of 40-mm convex on the left side and 42-mm convex on the right side, the other with both radii identical at 41 mm. The two lenses clearly look identical. Which has the nonequal radii? The answer is that the element on the left has the nonequal radii and the element on the right is perfectly equiconvex. During the assembly operation, the technician or assembly person will typically look at the reflection from each side from an overhead light source, and the side with the smaller reflected virtual image is the shorter radius. Unfortunately, the reflected imagery will look virtually identical from these two radii. What should be done in this case?

Figure 25.6
Two Nearly Identical Elements



- The best thing to do is either make the radii equal (both convex or both concave with the same radii), or make them sufficiently different so as to easily determine the correct orientation.
- If you cannot make the radii the same, perhaps the best thing to do is to place an intentional bevel of a size or face width which can easily be distinguished on either S1 or S2. The reason this is a good idea is that the shop will most certainly place the bevel on the correct surface.
- Another approach is to request that the shop put an arrow pointing to the second surface showing the direction of light. This is a reasonable idea; however, you may be dealing with a shop whose practice is to put the marks following a different convention, and then you have a really serious problem!

Insufficient Sampling of Fields of View or Aperture

Computer programs, no matter how sophisticated, do only what they are told to do. If you specify, for example, semifields of view of 0, 7, and 10° off axis, the optimization algorithm will work specifically on those

fields of view, and all other fields will be totally ignored as if they were nonexistent. If you have a system with higher-order aberrations and/or aspherics, it is very likely that the performance may degrade at field positions between these fields. Thus, for example, you may ultimately experience poor performance at 3 to 4° off axis. This was found to be the situation in our double Gauss case study in Chap. 22.

An MWIR infrared system with at least one aspheric surface was initially designed over five equally spaced fields of view. Unfortunately, during the final testplate fit, only three fields were used during the optimization. When the lens was tested, it performed superbly on axis and at its full field. However, at intermediate field positions the performance failed its specifications miserably. It is important to assure that the fields of view are sampled sufficiently during all phases of the lens design optimization. It is wise to evaluate performance of the system over 5 to 10 equally spaced field positions. It should be noted that sampling in pupil space is also important and not to be ignored.

Images Upside Down or Rotated

In visual systems the imagery must be both erect and right handed. In systems used for imagery onto a CCD or similar sensor, the image inversion and/or handedness can often be taken care of in the electronics. The orientation of the image can be verified by tracing a nonsymmetrical object through the system, which was described in Chap. 8.

Some years ago we had a panoramic system which used a prism to scan a wide azimuth field of regard. This form of system introduces image rotation which must be canceled by using another rotating prism subassembly such as a Pechan prism. Just after the machine shop started manufacturing the first components for the prototype, one of us in the team decided to carefully check once again if the direction of rotation of the Pechan prism was correct. Indeed, our original design had the direction of a Pechan prism rotation set incorrectly. Luckily there were not a lot of parts machined yet. If we did not discover this mistake on time, it would have been a major disaster, since these types of systems are generally very expensive.

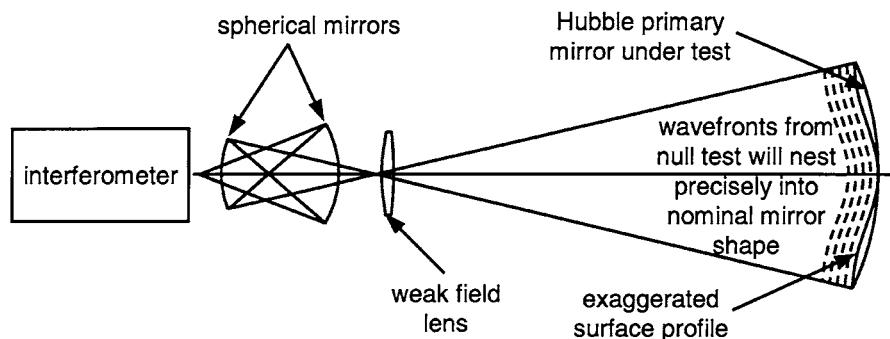
This type of panoramic system has to be designed so that the image is properly oriented in the nominal (zero angle scan position),

including the rotating element with a correctly determined axis of rotation. The second issue is the direction of rotation and the magnitude of angle of rotation. Generally, the magnitude of compensating element rotation is one-half of the scanning element angle of rotation. However, the direction of rotation has to be studied carefully in each specific case, since this is poorly covered in literature, and mistakes are very easily made.

The Hubble Telescope Null Lens Problem

As many of us know, a 1.3-mm error in the spacing of the null optics caused a significant error in the aspheric shape of the primary mirror of the Hubble telescope. The basic interferometric null test is shown in Fig. 25.7. An interferometer with a diverger lens similar to the system shown in Fig. 15.7 was used. Following the diverger focus, the diverging light is reflected from a concave spherical mirror, and after forming an intermediate image, it is reflected from a second concave spherical mirror which forms an image to the right of the first spherical mirror. A field lens is located at the last intermediate image as shown. The light then proceeds to the surface under test. The whole purpose of this test setup is to create a wavefront, which matches exactly the nominal mirror surface at the nominal location of the mirror surface. In other words, the null test creates a wavefront that precisely and perfectly nests into the nominal mirror surface under test. The extremely weak field

Figure 25.7
Basic Setup of
Hubble Telescope
Null Optics



lens appears to be doing nothing; however, since it is located at a highly aberrated image position, there is a significant difference between the paraxial rays and the real rays transmitting through the element, and this, in effect, allows the field lens to successfully balance the higher orders of spherical aberration. In the model we have developed, shown in Fig. 25.8, we have a residual double-pass optical path difference of 0.002 wave rms.

In initially setting up the test, it is imperative that the two mirrors and the field lens be properly positioned with respect to each other. In order to accomplish this, diverging light from the interferometer is first retroreflected from the right concave mirror back into the interferometer, as in Fig. 25.9a. When a null condition or straight fringes are seen, this establishes a precisely known spacing between the diverger focus and the mirror. The field lens now needs to be positioned, and a low-expansion invar metering rod of a precisely known length with slightly convex polished ends is now located so that the light from the interferometer focuses on the left-hand end of the rod and returns into the interferometer, as in Fig. 25.9b. Once again, when a null fringe or straight fringes are seen, we can be confident that the focus of the light is on the end of the metering rod. Now the field lens is just barely touched to the opposite end of the metering rod and bonded in place. The metering rod is now removed, and we are left with the right-hand

Figure 25.8
Design Similar to
Hubble Telescope
Null Test

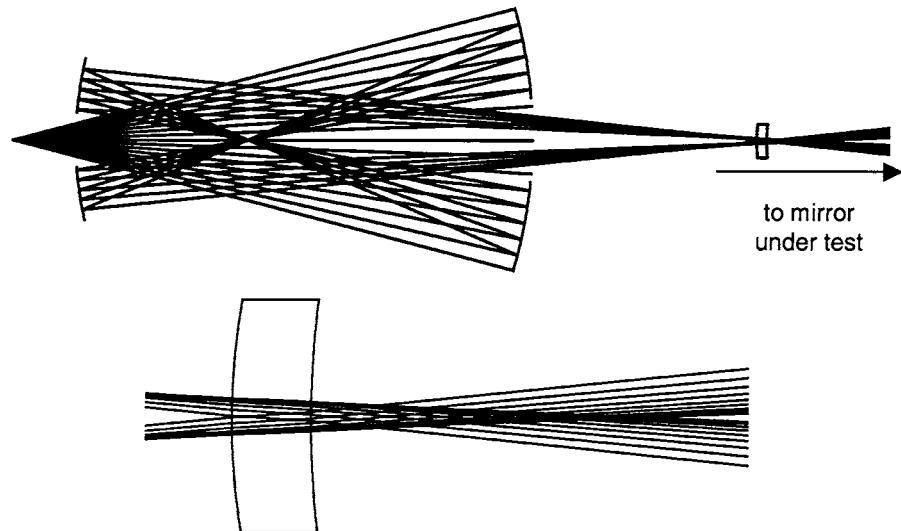
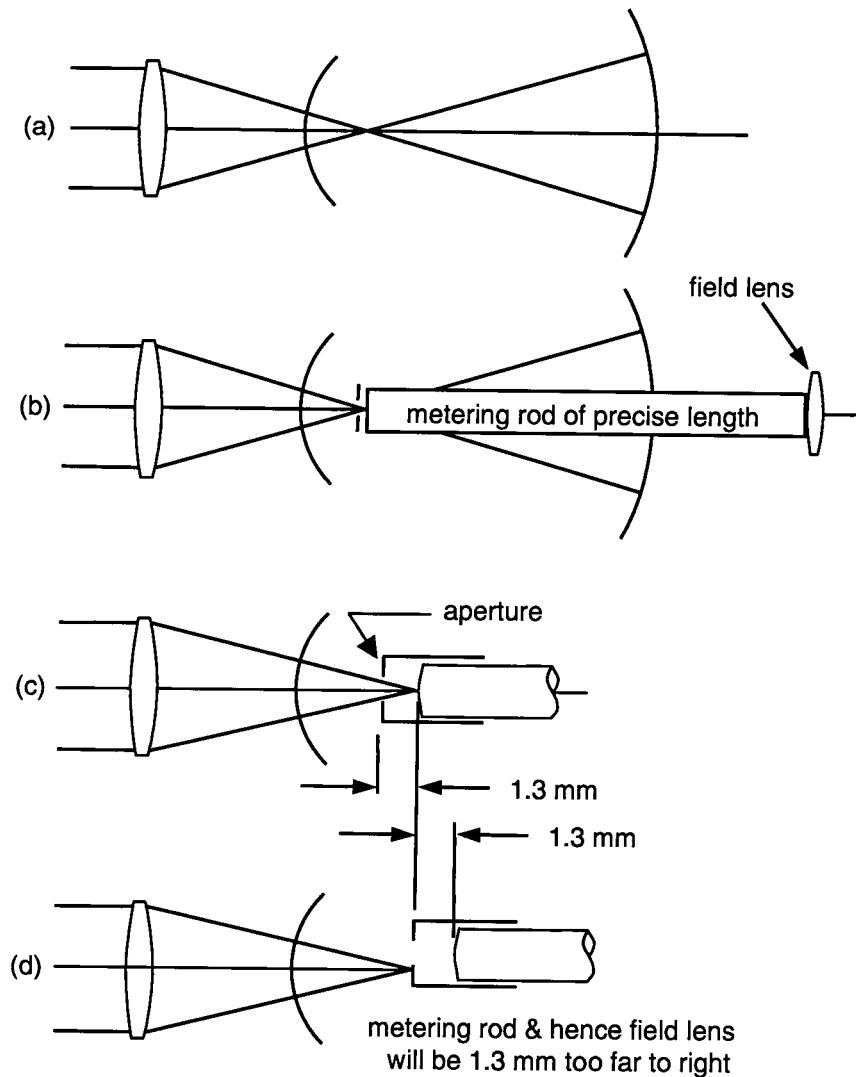


Figure 25.9

Illustration of the Hubble Telescope Null Lens Problem



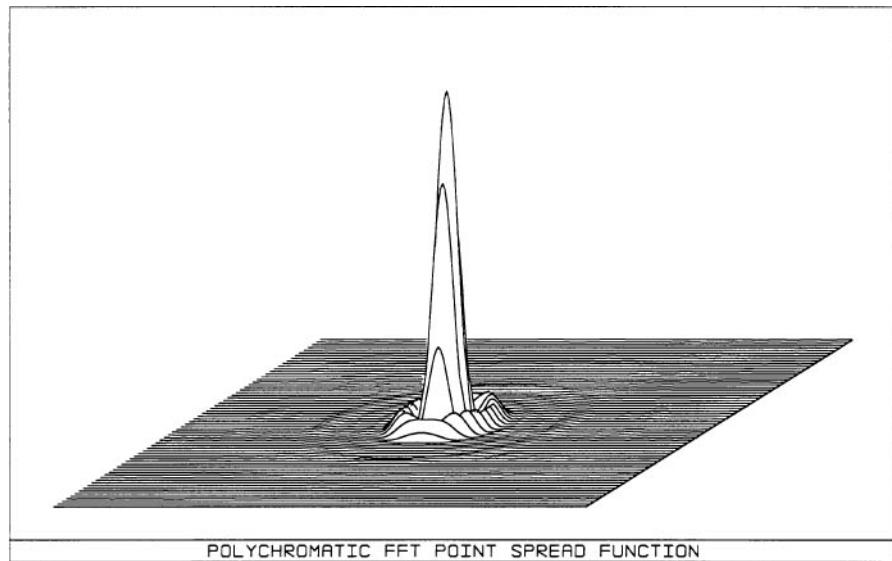
mirror and the field lens properly spaced with respect to each other. A further procedure is then used to locate the left-hand spherical mirror.

In order to assure that the light was properly incident onto the metering rod, a cap with a small aperture was placed onto the end of the rod, as shown in Fig. 25.9c. The cap was painted flat black so that if the system were misaligned and the laser light were to strike the cap

and not pass through the aperture, then no fringes would be seen in the interferometer. The aperture plane was approximately 1.3 mm from the end of the metering rod. After the system was initially tested, a piece of masking tape was placed over the aperture in the cap in order to keep dust out. Prior to testing, the tape was removed and a small piece of the flat black paint flaked off leaving the bare metal of the cap. During the setup of the null test components as outlined earlier, the diverger was inadvertently focused on the area of the cap where the paint had flaked off, and this caused the metering rod to be 1.3 mm too far to the right, as shown in Fig. 25.9d. This resulted in the field lens also being 1.3 mm to the right, and this was the problem. We would normally think that a bare metal surface like the top of a tin can would hardly be good enough to produce straight fringes in an interferometric test. However, remember that the diameter of the focused spot is in the order of $2 \mu\text{m}$, and the surface is quite likely to be good over this diameter.

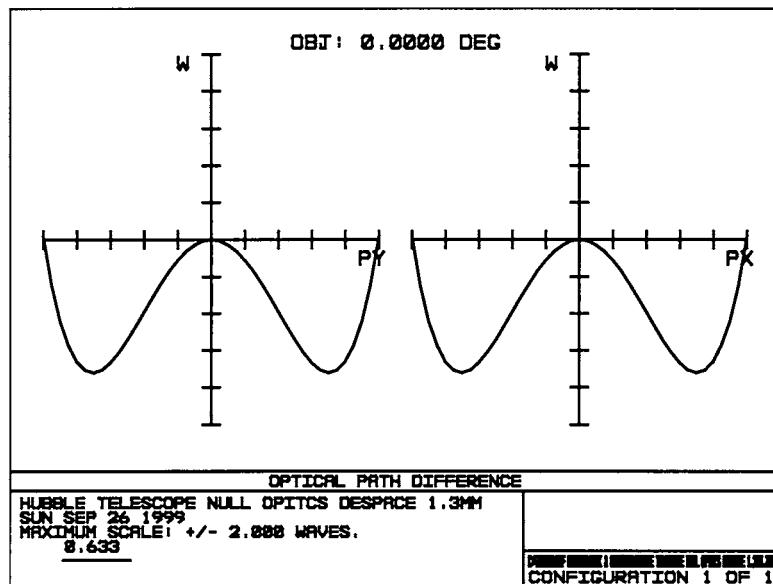
Figure 25.10 shows the image-point spread function for the nominal (perfect) null test. This is, in effect, representative of the imagery predicted for the telescope at its Cassegrain focus if everything were manufactured perfectly. Figure 25.11 shows a plot of the OPD for the null test with the field lens axially shifted by 1.3 mm, and Fig. 25.12 shows the

Figure 25.10
Image Point Spread Function of the Hubble Telescope Null Lens if Components Were Manufactured and Aligned Perfectly



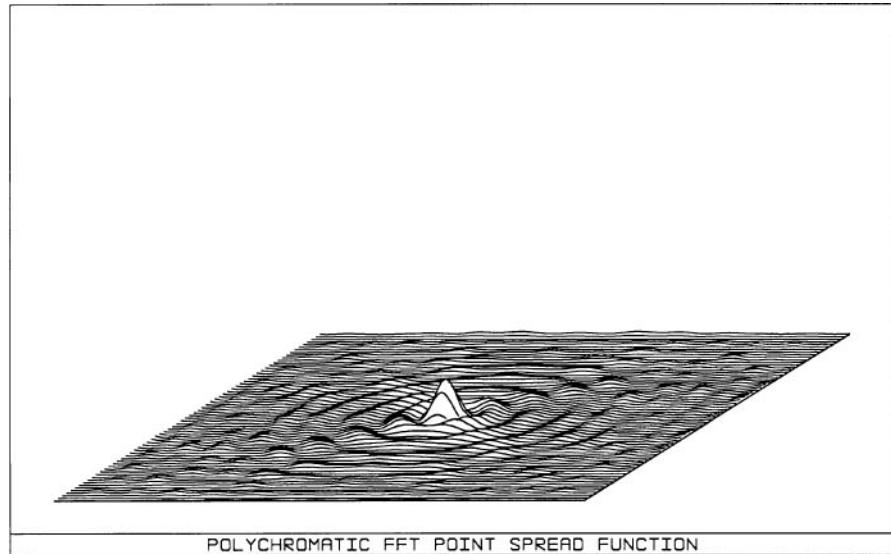
POLYCHROMATIC FFT POINT SPREAD FUNCTION

Figure 25.11
Optical Path Difference with Field Lens
Despaced 1.3 mm



resulting point spread function. These data are representative of what the imagery would be like for the system with the primary mirror manufactured with the null test including the shifted field lens. The Hubble telescope was ultimately repaired and was a great success as evidenced by spectacular imagery.

Figure 25.12
Image Point Spread
Function with Field
Lens Despaced
1.3 mm



Wrong Glass Type in a Precision Lens System

One of the worst fears to any designer or manufacturer is that an incorrect glass type is used for one or more elements. By the time the error is discovered, the system has likely been assembled and tested, and likely performs very poorly, even though all dimensions may have been met. Occasionally this happens as evidenced below.

A large manufacturer of precision optical glass provided 10 cut lens blanks of an 11-element system for a lens to be used by the U.S. Navy, and valued at over \$12M for the entire program. The lens fabrication shop called many months after the glass had been delivered, indicating that once the lens elements were fabricated and the system assembled the system “flat out did not work” and that their contract with their customer was in jeopardy of being canceled. They asked the glass manufacturer to reconfirm the precision index measurement data on the elements to ensure that they had received exactly the glass types they had ordered. Very precise prisms were manufactured from a set of blanks and a full compliment of the most accurate precision index measurements on the samples conducted. The measurements on the 10 elements were a mirror image of the original measurements in every detail, down to the sixth decimal place. It was also found that the supplier of the 11th blank had used the wrong glass type, which was not even close to the required equivalent glass type, and this is why the system did not work. By that time, the correct glass types were in stock, and were able to be provided in the requested quantity, and the rest of the order was successfully completed, with no more problems.

Single Use Camera with a Diffractive Achromat

A camera was developed by a large company that used a hybrid singlet meniscus lens, with the diffractive surface at the rear surface for achromatization. The diffractive surface was continuous, namely a non-binary kinoform, made with diamond turned molds, so it achieved 99.7% diffraction efficiency.

Extensive testing was done to verify performance and find any artifacts. The technical team concluded that a baffle of about 1.5 mm long will eliminate an artifact which was due to direct sunlight scattering especially around glint points in the image on windows and cars. The camera optics was thus designed with a 2 mm baffle.

The camera industrial designers had in mind what they called the “soap bar” look, and they did not like the “long” baffle. They shortened it to about 0.6 mm. This was done at a late stage of the project, quite close to production start. When the optical team found about the shortened baffle, they protested but were ignored.

Accentuated flare light showed up in a number of images taken with the industrial models made with a short baffle. This was caused by sunlight reflecting off vertical faces of the diffractive facets into the camera housing. It affected the upper left and right corners of the field. It was not apparent immediately what happened. The assumption that it was related to issues of diffraction efficiencies of the diffractive element turned out to be wrong. The suggestion by the technical team to lengthen the baffle to about 1.5 mm was considered to have a negative marketing impact from the point of view of the camera look and the possible delay in market introduction, so they wouldn’t extend it.

As a result, the project stopped. It never went to market trials, and never made it to customers. The marketing department essentially dismissed the technology and did not revisit it seriously after that, even though the root cause and the technical fix were identified.

There was later a digital camera that incorporated a diffractive achromat that went into production and the lens worked well. The product didn’t become a large volume seller due to nonoptical reasons and the company dropped all digital cameras after that.

Other companies continued in this technical area. One notable result was the successful use of diffractives in high-end zoom lenses.

Wrong Image Handedness

The design of a panoramic aerial reconnaissance camera involving two fold mirrors incorporated into the design for mechanical packaging reasons was nearly completed before someone checked to make sure the image on the film emulsion would be oriented properly when passing

through the photo interpreter's viewer. Sure enough, it was found to be right side up but reversed left for right, that is, "reverted." It was far too late and would be extremely costly to add or remove one mirror. Then someone realized that the image orientation would be correct if viewed through the back of the film rather than with emulsion out, as is normally the case. The user agreed to make this simple change in procedure and the project was saved with only minor embarrassment to all parties.

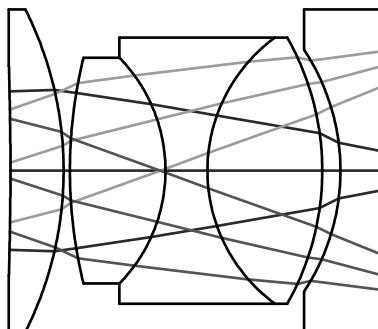
Cemented Triplet as Part of an Imaging System

A lens was designed with a steeply curved cemented triplet as shown in Fig. 25.13. A number of problems were encountered:

First, the crown—flint Abbe number difference of 17 yielded very short and steep radii. This geometry in turn created a highly asymmetric illumination of the UV radiation used for curing the cement, which in turn made the triplet more sensitive to thermal soak. Further, the steep radii also introduced more aberrations than desired.

The near-zero edge thickness on the 2nd positive element was difficult to manufacture. In addition, a steep aspheric (elsewhere in the design) had a sag profile that became quite ill-behaved just outside the element edge...this created problems with the diamond grinding machine.

Figure 25.13
Cemented Triplet
with Several Issues
Affecting Manufacturability



Several messages became clear:

- If you plan to use a cemented triplet in your design, discuss it with your manufacturer early in the design process to assure its producibility.
- Where possible, use larger Abbe number differences for crown and flint elements so as to weaken the individual radii.
- Avoid where possible aspherics, especially strong ones, and if you need aspherics, assure the surface is “well behaved” both within its clear aperture as well as just outside the element edge.
- It is important to consider the match of the coefficient of thermal expansion (CTE) of the cemented glasses as well as the housing over the required temperature range and shock conditions as appropriate.

All of the above factors must be carefully monitored during design to assure manufacturability of your lens.

Total Internal Reflection in a Cube Beamsplitter

A cemented dichroic cube beamsplitter was designed into a system as shown in Fig. 25.13 in order to image two separated wavelength bands to two sensors. For packaging reasons (to lengthen the optical path) a high index glass was used for the two right angle prisms comprising the cube. The lens design effort was done using a single block of glass representing the two cemented prisms. In most situations this is fine.

The “critical angle” is the angle of incidence normal to the surface in the denser material when the angle of refraction is 90° . At angles of incidence greater than the critical angle, the rays will totally internally reflect and the surface behaves as a mirror. The critical angle for the cemented cube is $\sim 55.2^\circ$ as shown in Fig. 25.14a. Figure 25.14b shows how ray 1 passes through the beamsplitter cube laterally displacing downward for the exaggerated cement thickness shown. Ray 2 exits the cube parallel to how it went in, ray 3 is at the critical angle, and ray 4 TIRs.

The problem was that a noticeable sector of the pupil totally internally reflected, and for the application intended this was unacceptable. High-index fluids were tried in the cement space, but this did not work.

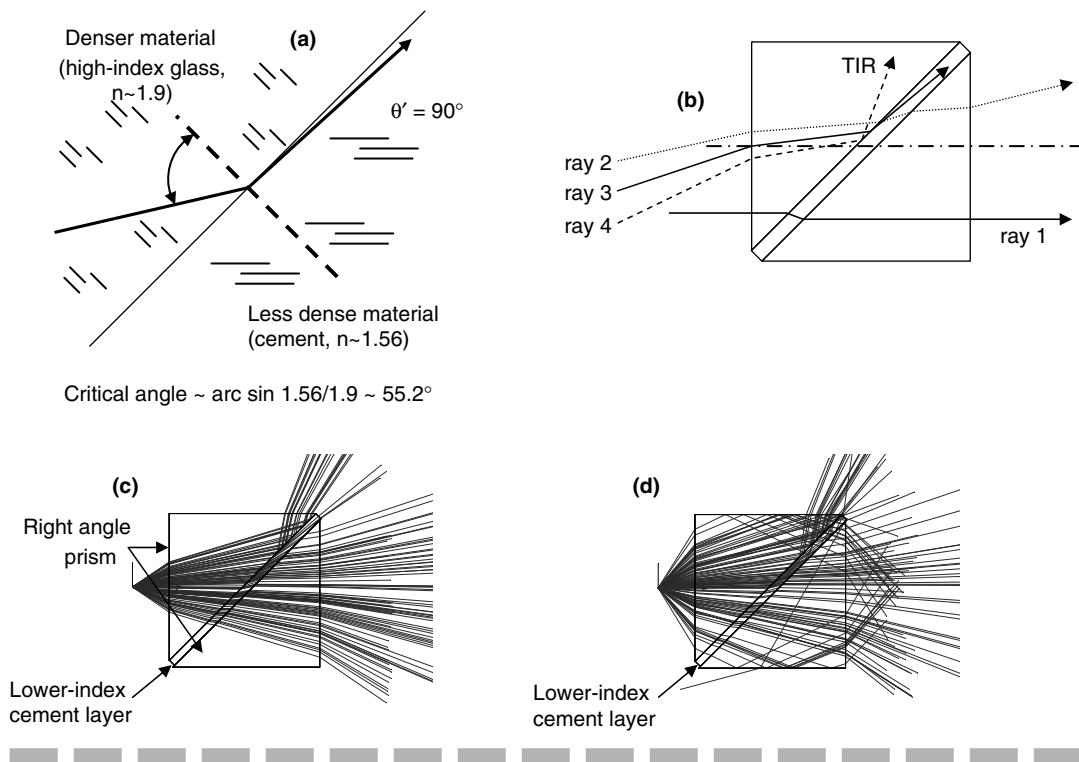


Figure 25.14
Total Internal Reflection (TIR) in a Cemented Cube Beamsplitter

Nonsequential ray tracing can show dramatically what is happening. Figure 25.14c shows an $f/0.9$ object cone, and it is clear just where the rays TIR. In Fig. 25.14d the object cone angle is increased, and other potential problems become evident. Fortunately, the limiting cone angle is often set by appropriate apertures in the system and the extreme ray angles as shown here are not present. Do keep in mind that you must consider the field of view, as there will be different ray angles at each field of view.

It is imperative to assure that your design does not have total internal reflection issues as described above. TIRs can occur also in steep angles of incidence in lens systems with steeply curved surfaces, especially when you have a large refractive index difference between the glass and cement.

Diffractive Optics Issues

Diffraction Efficiency

The notion of diffraction efficiency in diffractive optics should be used with great caution, especially when describing how well a diffractive works (or is expected to work), for a specific application.

Usually, diffraction efficiency is defined as the ratio of the amount of light in the desired diffraction order (fundamental positive order for example) by the amount of light falling onto the diffractive element.

This is a theoretical definition which applies very seldom to real life, for the following reasons:

- The real efficiency of a diffractive element is very seldom equal to the theoretical value which can be calculated by either scalar or vectorial methods; most of the time the actual value reduced due to systematic fabrication errors (etch or groove depth errors, lateral misregistrations in multilevel fabrication, resolution limits in lithography, etc.).
- Efficiency reduction occurs also when using scalar theory of diffraction to predict efficiency while the structures in the diffractive are close to the wavelength, where scalar theory is much less accurate. However, predicting efficiency with vectoriel methods for anything else than a linear grating is usually a difficult task.
- Most of the time, these efficiency reductions push more light into the zero order (the central spot in far field reconstructions, or uniform noise in near field reconstructions). With severe fabrication errors, especially due to lateral misregistrations in optical lithography, higher parasitic diffraction orders can steal some precious light from fundamental orders.
- On the other hand, the real efficiency can also be substantially larger than the theoretical predictions (scalar or vectoriel). This is usually a good surprise to the optical engineer, and happens often when designing binary Fourier pattern generators, binary spot array generators or binary far field beam shapers, where efficiency is actually doubled from 40% to 80% by the overlapping of both fundamental orders.

- Additional factors reducing the efficiency although to a lesser level, are Fresnel reflections from the structured and/or the flat surfaces of the diffractive as well as material absorptions which do not account in the traditional efficiency calculations.

In a general way, when calculating the efficiency of a diffractive element for a specific application, it is better to redefine it as the amount of light accounting for the desired functionality, rather than considering diffraction orders.

Diffractive Lenses versus Refractive Lenses

It is a common misconception that one can replace a refractive lens with a diffractive lens, for example to gain space and weight or reduce fabrication costs (by mass replication in plastic).

There are four main reasons to this:

- Efficiency reductions (refractives are 100% efficient for any wavelength), even for the optimal design wavelength.
- Chromatic aberrations (which are much stronger in diffractives than in refractives, and have opposite sign), in transmission as well as in reflection modes (refractives have no chromatic aberrations in reflection mode).
- Polarization effects in diffractives not occurring in refractives (especially when diffraction angles are large).
- High launch angle problems in diffractives (shadowing effect due to microstructures, etc.) not occurring in refractives.

Diffractive lenses are best used in the following cases:

- As standalone elements for monochromatic illumination under angles lower than 45°, with maximum deflection power of 25°.
- As hybrid elements combined with refractives, to produce achromatic and/or athermal elements.
- As large arrays of elements with 100% fill factors along with cheap mass fabrication costs.

The following example shows how one can make a very bad and a very good choice by inserting a diffractive element into an OPU (Optical Pick up Unit) for CD-DVD drive.

VERY BAD CHOICE VERSION Let's aim to replace the DVD-ROM OPU refractive lens by a diffractive lens with the exact same prescriptions (because it will be lighter, cheaper, and smaller than its refractive counterpart).

The NA of the DVD OPU lens is 0.60 for 650 nm, and the overcoat of the DVD media is 600 μm . Any optical design software would easily generate an aspherical phase polynomial over a plane surface to describe the diffractive, and correct for spherical aberrations. After fabricating the diffractive and replacing the refractive lens, the OPU does not work, WHY?

- Although we use laser illumination, the efficiency of the diffractive lens drops sharply at two edges of the lens because of severe polarization effects when the grooves of the lens are aligned to the polarization direction of the laser beam (because of the high NA), and thus the effective aperture of the lens becomes strongly elliptical rather than circular, reducing throughput, creating noise (zero order), and changing the spot size and geometry on the quad detector for track/focus control.
- Additional severe coma is appearing when the OPU tilts mechanically the lens by a maximum of $\pm 3^\circ$ to follow DVD tracks, which is not appearing in the refractive lens (with exact same prescription), simply because its surface is curved rather than planar (minimizing coma).
- Finally, diffraction efficiency for this application is not even close to the theoretical amount of light in the fundamental order (focusing into the spot) first because scalar theory is only partially valid here (because of high NA), and second because the amount of light desired here is the light falling within the first ring of the Airy disk, which is maximum for a strong Strehl ratio, and not the whole light in the fundamental order, spread over all the sides lobes.

This is why pure diffractive lenses are not integrated today in CD/DVD drives.

VERY GOOD CHOICE VERSION Let's try to replace the two different CD and DVD OPU lenses with a single hybrid lens which would focus two different spots (0.45 NA and 0.60 NA) for two different wavelengths (780 nm and 650 nm) correcting two different spherical aberrations (for 1.2 mm overcoat and 600 μm overcoat) in order to read the CD media and the DVD media without changing mechanically the lens.

The solution to this problem could be a hybrid refractive/diffractive lens, with a strong first spherical profile and a second aspherical profile on top of which a diffractive profile is fabricated, tuned at only 50% efficiency. The CD spot would be generated by two surfaces (the two refractives and the 50% undiffracted light remaining in the zero order of the diffractive) and the DVD spot by the combination of three surfaces (the two refractives and the 50% diffracted light by the diffractive). This approach has been patented by Matsushita in 1995 and is implemented in most CD/DVD drives today.

This typical practical example shows that the best choice is to use a hybrid refractive/diffractive lens in order to produce a functionality which could not be produced by neither a pure diffractive nor a pure refractive lens or lens compound.

Case of the Miscoated Mangin

A catadioptric imager (reflective and refractive components) was designed operating over the 3–5 μm spectral band. The objective was in a Cassegrain arrangement, but consisted of a spherical primary mirror (with a central hole) and a Silicon Mangin secondary mirror to correct the spherical aberration. The Cassegrain focal point was approximately 200 mm behind the vertex of the primary mirror (Fig. 25.15a). In the actual system, light entered a barrel containing relay elements before reaching this focus. The relay barrel was held by a threaded collar that fit inside the hole of the primary mirror.

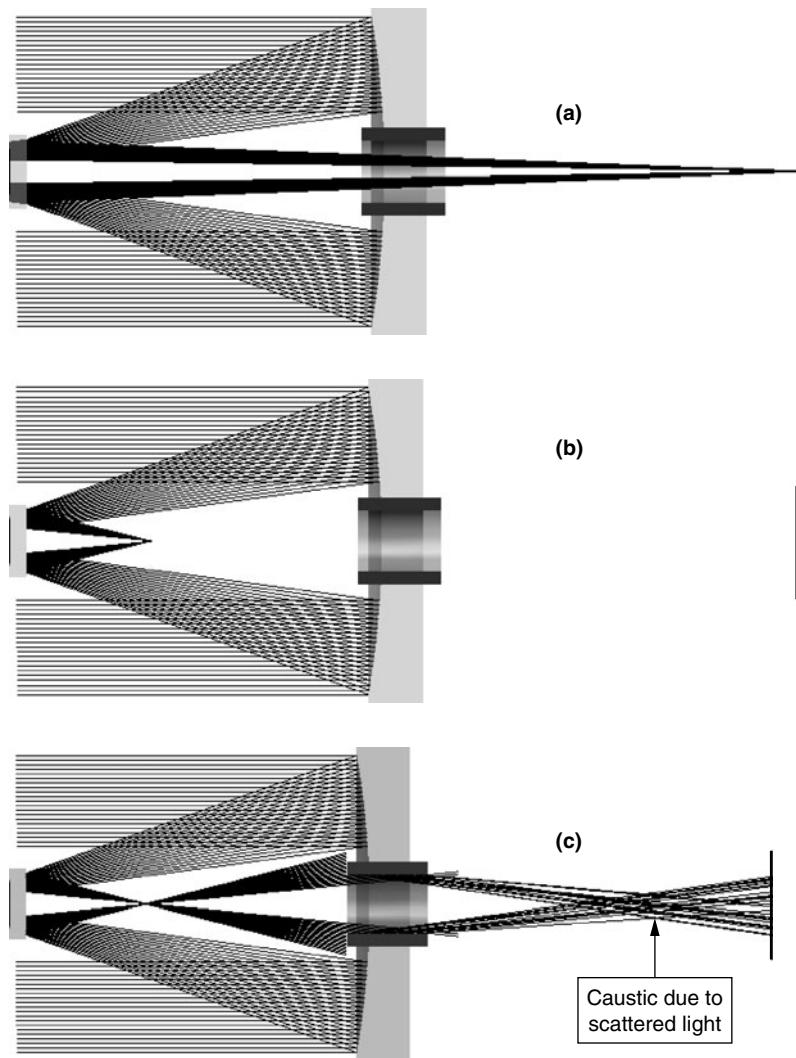
Initial testing of the full system (objective and relay) with a 3–5 μm MTF bench indicated virtually no resolution whatsoever. An initial examination of the hardware did not indicate anything obviously wrong, so the individual subcomponents were evaluated.

First, the Cassegrain objective was tested alone. Using the 3–5 μm MTF bench, there was no focus at the design location 200 mm behind the vertex of the primary mirror. A significantly aberrated focus (more aberrated than the computer prediction) was found about 160 mm after the primary.

An initial suspicion was that the coatings on the Mangin mirror may have been reversed. The 3–5 μm antireflection coating on the front of the Mangin was expected to be highly reflective in the visible, so there was no visual method to determine if the coatings were reversed. Since

Figure 25.15

- (a) Nominal Objective,
- (b) Objective with Miscoated Mangin,
- (c) Objective with Miscoated Mangin and Scattered Light Focus



the Mangin was already mounted and aligned, it was not desirable to disassemble it for testing. Instead, a computer simulation of the system found that if the front surface of the Mangin was reflective, the focus would fall well in front of the primary mirror. Since the focus found with the MTF bench was about 160 mm after the primary mirror, it did not appear that this could be the problem.

Various tests were performed over the next several days simulating possible scenarios (for example, what would be the effect if AMTIR was accidentally used instead of Silicon for the Mangin?), but no other failure mechanism was identified. At one point, with the lights off in the lab, a faint visible caustic was noticed at nearly the same location as the previously discovered, severely aberrated, 3–5 μm focus. The 3–5 μm source at the focus of the collimator was replaced with a bright visible light and, as it should, the visible light reflected off the front surface of the Mangin and formed a focus before the primary mirror (Fig. 25.15b). However, the light continued to diverge toward the primary, and it illuminated the threaded collar that normally held the relay optics barrel. The cylindrical, threaded barrel acted as an axicon (a cone shaped mirror) and formed a severely aberrated visible caustic about 160 mm after the primary, the exact same location as the severely aberrated focus for the 3–5 μm system (Fig. 25.15c).

The Mangin was dismounted and sent to the coating lab to be tested. The coatings were indeed reversed, with the 3–5 μm reflective coating having been applied to the front surface of the Mangin.

Telescopes and Polarization

The 2.1 m telescope at Kitt Peak National Observatory, near Tucson, AZ, is an equatorial mounted Ritchey-Chretian. There are two foci: a classical Cassegrain, and a Coude. The Coude works at $f/34$ and uses a total of five mirrors to bring the light down the equatorial axis. In addition to the Cassegrain primary and secondary mirrors there is a third mirror that directs the Coude beam into the declination axis. Another mirror, in the declination axis, sends the beam toward the equatorial axis, and the fifth and final mirror makes the beam coaxial with the equatorial axis.

The Coude beam rotates as the telescope is pointed to different locations in the sky. In the late 1960s the observatory installed an optical derotator to stabilize the orientation of the Coude image. An all-reflective “K-mirror” was used as the derotator. A mirror derotator was used to provide the widest possible spectral range of transmission.

In service, the derotator transmitted well when the telescope was in certain positions—and transmitted nothing when the telescope moved to a new position! The Coude path of the telescope used two oblique mirror reflections, which polarized the light beam. Similarly, the “K-mirror”

used oblique reflections, which produced polarization. The K-mirror acted, therefore, as a polarizing analyzer, and would cut-off transmission due to polarization from the telescope Coude path.

The K-mirror was eventually replaced with a prism assembly using total internal reflection, and the problem went away—although not without considerable embarrassment on the part of some very senior optical designers and astronomers!

Rule of Thumb and Hints

General Optical Design Topics

There are many “rules of thumb” in optics, and we will summarize the most useful ones here. While they may be discussed elsewhere in this book, this chapter is a compendium of the most important and useful rules of thumb and hints.

- Diffraction-limited Airy disk diameter = $2.44\lambda f/\#$ in units of λ .
 - In the visible the Airy disk diameter is approximately equal to the $f/\#$ expressed in micrometers.
- Diffraction-limited angular Airy disk diameter = $2.44\lambda/D$ rads, where D is the entrance pupil diameter in the same units as λ .
- Resolution of a diffraction-limited system in the visible is approximately $136/D$ in arc seconds, where D is the diameter of the entrance pupil in millimeters.
- A system will provide image quality, which is nearly indistinguishable from perfect, if the optical path difference from a nearest reference spherical wavefront reaching the image departs from sphericity by one-quarter of the wavelength of the light or radiation. This is the Rayleigh Criteria.
- The depth of focus for one-quarter-wave peak-to-valley optical path difference = $\pm(2\lambda f/\#)_2$ in units of λ .
 - In the visible, the depth of focus is approximately equal to the $(f/\#)_2$ in micrometers.

- The spatial frequency where the MTF goes to zero = $1/(\lambda f/\#)$ = 1000 line pairs/mm for an $f/2$ lens at a wavelength of $0.5 \mu\text{m}$.
- The Nyquist frequency is the highest spatial frequency that a pixelated sensor can successfully record. It is $1/(2 \text{ pixel period})$ expressed in line pair per millimeter.
- The MWIR wavelength (3 to $5 \mu\text{m}$) is approximately 8 times the visible and the LWIR wavelength (8 to $12 \mu\text{m}$) is approximately 20 times the visible.
- RMS wavefront error is approximately $1/5$ to $1/3.5$ of the peak-to-valley optical path difference.
- “Clip it at the bud”—correct aberrations as close to where they are introduced in your system as possible. This will be easier to do and yield better overall performance.
- Always consider the possible effects of stray light, and provide suitable baffles and low-reflectivity interior system finish to attenuate it.
- Always consider the possible effect of ghost images due to multiple reflections from lens surfaces in your system. While ghost images are, for the most part, not a problem, it is best to analyze the situation.
- Make sure you tolerance your optical system and use realistic tolerances, otherwise your system will not be producible and/or will be very costly.
- Tolerances do not lie!
- The effect of tolerances is somewhat like standing at the edge of the Grand Canyon (well, not really, but this analogy is thought provoking). If someone pushes you a little on the shoulder, you will be fine. However, if they push very hard, you will tumble in rather fast. Tolerances are similar; small tolerances are generally fine, but as soon as they increase beyond a certain critical level, the performance will get bad very quickly.
- If your goal is to provide near to diffraction-limited performance, as your tolerances and the errors introduced by them increase as you error budget your system, the predicted MTF will degrade slowly at first and then faster and faster.
- If possible, try and assure that your specifications are based on what is *functionally* required for your system.

- Talk with your lens manufacturer to assure that the lens elements and other optical components are producible and your tolerances are reasonable. The same holds for the mechanics.
- Avoid very small, nearly concentric airspaces. This may lead to tight tolerances. You may be able to cement the two elements, which will also eliminate two antireflection coatings.
- Wherever possible, use good glasses which are easy to manufacture and low in cost. While the use of anomalous dispersion and other nonconventional glasses are sometimes valuable, they are not always required.
- Do not use aspheric surfaces unless they are mandatory, and if you do use them, make sure they are producible.
- The best rule of thumb regarding element thickness seems to be "if it looks good, it probably is good."
- Avoid elements with very thin edges and also very thin center thicknesses. They may warp during manufacturing, be costly, and/or be difficult to mount.
- For low-order aspherics use either a conic or a fourth-order asphericity, not both. While you can use both a y^4 and a conic, they often have a similar effect and can "beat" against one another.
 - If your surface is reasonably curved, a conic can be used.
 - If your surface is nearly flat, use only the fourth-order asphericity.
 - When you do use aspherics, work with the lower orders of asphericity first, and only use as high an order as you really need.
- Only use diffractive surfaces if you really need them.
 - Assure through discussions with your manufacturer that the diffractive profiles are producible.
 - You should be very concerned about diffraction efficiency and scattering, especially in the visible or at lower wavelengths.
 - If you use binary optics, make sure you have sufficient phase steps to minimize scattering.
- If possible, use only good quality glasses, which have low sensitivity to stain, bubbles, and other parameters.
 - Check the cost and availability of the glasses.
 - Avoid glasses with any parameters that are nonstandard such as transformation temperature and stain characteristics.

- Pay attention to possible polarization thin-film coating issues.
- If your system is to be used in a polarized light environment, stress birefringence should be considered.
- Make sure your image is right side up and right handed, especially in visible systems.
- Always try to match 100% of the radii to existing vendors' testplates.
- Remember, scattering increases with decreasing wavelength. You will thus have significantly more scattering in the blue and UV portions of the spectrum.
- When evaluating the effects of stray light and scattering in visible or IR systems (UV too), put your eye figuratively at the sensor and look outward and ask yourself "what do you see." The answer will be very revealing with respect to stray light and scattering, as well as to the means for controlling the stray light.
- When working with asymmetrical systems in which you have small tilted surfaces or decentered elements, you can validate your setup and the sign convention by increasing the tilt or decentration to a very large value so it shows up clearly on a layout.
- Make sure you consider the thermal and other environmental requirements for your system. If athermalization is required, do not wait until the last minute to determine how to accomplish the task.
- Be extremely careful in working with and interpreting the terminology and conventions used in lens design programs. A good example is the use of the term *spot size*; is this a spot *diameter* or a spot *radius*?
- Finally, remember that hindsight is diffraction limited!

Optomechanical Topics

- Check image orientation early in the design.
- Mechanical and optical characteristics of materials are not constant with temperature.
- Optical surface deformations usually are more important than stresses in optics.

- Don't forget to consider the effects of the environment.
- Tolerances should be tight wherever the cost penalty is small and performance gain large.
- A very few adjustments made at final assembly can optimize system performance.
- Too many adjustments are as bad as none at all.
- Sealing the instrument will reduce entry of moisture and other contaminants.
- Preload applied to an optic in all three directions should equal that optic's weight multiplied by the greatest anticipated acceleration.
- Preload provided by a threaded retainer is very approximately five times the applied torque divided by thread pitch diameter.
- Tensile stress in any optic should not exceed 6.89 MPa (1000 lb/in²)
- Flat springs used to preload optics should have spherical or cylindrical pads as interfaces to glass.
- Interfaces between convex lens surfaces and mechanical surfaces should be conical.
- The interfaces between concave lens surfaces and mechanical surfaces should be toroidal with the toroid radius at least 0.5 times the lens radius.
- Tolerance on a toroidal radius can be very loose.
- Spherical interfaces are usually not worth their cost.
- Flat bevels used as mechanical references should be adequately normal to the optical axis.
- Axial and radial compliance built into lens and mirror mountings minimizes stress at extreme temperatures.
- Safety factor on bond area should be at least four.
- Diameter-to-thickness ratio of a mirror bonded on its rear surface should be at least 6:1.
- Adhesive bonds on multiple-component prisms should be placed on only one prism—usually the largest.

- Torsional flexures are better than bearings for small motions.
- Always document the design for future reference.

Diffractive Optics

- When should I use vector (or rigorous EM—electro-magnetic) rather than scalar theory of diffraction to predict diffraction efficiency for a given diffractive element?
 - If minimum feature sizes in the diffractive are at least four or five times larger than the minimum period in the diffractive, use scalar theory. If lower, use local grating approximation combined with vector theory.
 - In any case, reconstruction angles are the same for either scalar or vector theories.
- What is the smallest feature size for a given diffractive lens?
 - Smallest feature is usually the smallest fringe period. Smallest period in a diffractive lens is wavelength/NA.
 - If the lens is a binary lens (that is, 2 phase levels), this minimum feature becomes half the period, and if the lens is a 16 phase levels element, this smallest feature becomes a 16th of the smallest period, which shows how fast the resolution limit of the fabrication tool can be reached by trying to fabricate a diffractive lens with large number of phase levels (see also next rule of thumb).
- What is the maximum number of phase levels I should choose when using conventional optical lithography to fabricate diffractive optics?
 - If efficiency is critical and budget is of no importance, go up to 16 levels for 98% theoretical efficiency and about 85 to 90% practical efficiency. It makes no sense to go over 16 phase levels, since the slight increase of efficiency would be completely washed out by the successive fabrication errors (etch depth error and lateral misregistrations), and fabrication costs increase.
 - If fabrication budget is tight and efficiency should still be maximum, use 4 or 8 phase levels maximum, to yield a practical efficiency of 70 or 80%.

- If budget is very tight and reconstruction is in far field, use a symmetrical reconstruction with only 2 levels (binary elements) in order to produce 70% or more efficiency (overlapping of many orders, especially both fundamental orders). For nonsymmetrical far-field reconstructions or near-field reconstructions, expect only 35% efficiency for binary elements.
- In summary, when considering the number of phase levels, in terms of optical efficiency, 4 is very far away from 2 (binary element), and 8 is very close to infinity (analog surface relief element).
- When should I use diamond tip machining rather than optical lithography to fabricate diffractive optics?
 - Use diamond turning for on-axis blazed Fresnel lenses, blazed cylindrical lenses, blazed linear gratings, or hybrid optics.
 - Use optical lithography for arrays of Fresnel lenses, beam shapers, CGHs, off-axis Fresnel lenses, and any other binary optical element.

This page intentionally left blank

GLOSSARY

Abbe number (also *V* number) $(n_d - 1)/(n_F - n_C)$, where n_d is the refractive index at wavelengths $d = 0.5876 \mu\text{m}$, $F = 0.4861 \mu\text{m}$, and $C = 0.6563 \mu\text{m}$. Also called *V* number.

aberration Geometrical errors in imagery whereby a perfect (or stigmatic) image is not formed. Typical aberrations include spherical aberration, astigmatism, coma, and chromatic aberration. The aberration of distortion does not affect the image quality, but rather the image position. Similarly, field curvature creates an image on a curved image surface. Lens bendings, locations, powers, glass types, as well as number of lenses and stop position, are all used to minimize aberrations.

achromatic lens A lens using two or more glass types which brings two colors to a common focus. Refractive along with diffractive optics can also lead to achromatic designs.

afocal lens A lens system which takes collimated light input and produces collimated light out, such as a pair of binoculars, a telescope, and a beam expander.

Airy disk The central maximum of the diffraction pattern from a perfect optical system with a circular unobscured aperture. The diameter of the Airy disk is $2.44 \lambda f/\text{number}$.

anomaly, image False or “ghostlike” images in thermal infrared systems which are caused by the detector “seeing,” or sensing energy from, portions of the system interior (rather than scene energy). If the system interior is at a different temperature and emissivity than the scene, bright (or dark) areas will appear on the display. Sometimes this can be serious enough to make the system nonfunctional.

aperture stop The location within a lens system where the chief, or principal or central, ray passes through and crosses the optical axis. This is the location within the lens where the ray bundles appear to pivot about. The presence of a mechanical limiting aperture typically creates the limiting size.

aplanatic lens A lens which is free of third-order spherical aberration and coma.

apoachromatic lens A lens in which three colors have been brought to a common focus.

apparent field of view The field of view that the eye sees when looking through an eyepiece. The ratio of the apparent field of view to the real field of view in object space is the “magnification.”

aspheric surface A lens or mirror surface which departs from a spherical shape. Conic surfaces (paraboloidal, hyperboloidal, etc.) as well as higher-order aspheric departures are often required for aberration reduction.

astigmatism An aberration in which light in one plane (the “plane of the paper” or meridional plane) focuses at a different location from the orthogonal plane. Astigmatism varies in proportion to the aperture and quadratic with field of view.

axial color The aberration whereby different colors focus at different distances from a lens. Primary axial color is the residual between the upper and lower wavelengths. Secondary axial color is the residual between the upper/lower wavelengths and the central wavelength.

back focal distance The distance from the last lens vertex to the image.

binary optics Diffractive optics where a staircase approximation to a kinoform is used for the surface profile.

binary optics (or digital optics) Micro-optical elements fabricated by the means of binary chrome masks, optical lithography, and etching (similar to binary- or digital-microelectronics).

birefringence Having two refractive indices, as in some crystals.

blocking, lens A support whereby many spherical lenses can be mounted and optically ground and polished at one time.

boresight error An error of alignment of the optical axis of two related systems parallel to each other. It is also an error of parallelism of an optical and a related mechanical axis. This is generally expressed as an angle.

Cassegrain telescope A reflecting system consisting of a concave primary mirror and a convex secondary mirror. The image is located behind the vertex of the primary. The mirrors are typically aspheric in shape (paraboloidal/hyperboloidal) for the “classical Cassegrain,” and both hyperbolic for the coma-free Ritchey–Chrétien Cassegrain.

catadioptric system An optical system consisting of both lenses and mirrors with optical power.

chief ray The ray passing through the center of the aperture stop of a lens or a mirror system. Also called a “principal” or “central” ray.

chromatic aberration An axial or off-axis aberration whereby different colors have different focus positions, magnifications, spherical aberration, or other aberration forms.

clear aperture The element diameter required for complete imagery over the full field of view. The term “optical clear aperture” refers to the lens (or mirror) aperture and the term “mechanical clear aperture” refers to the aperture created by mechanical features.

clipping The effect in an infrared optical system whereby the beam is vignetted or clipped by an aperture within the system. This often produces undesirable cosmetic effects on to the imagery.

cold finger A cooling device mounted directly behind a detector in thermal infrared cameras.

cold shield A cold aperture within a dewar in an infrared system inside of which there is image-forming radiation. Radiation outside of the imaging cones and inside of the cold shield can be seen by the detector and is typically interior system structures.

cold stop A cold aperture within a dewar in an infrared system which is also the aperture stop of the system. It allows for the detector to only see scene energy and no system interior.

cold-stop efficiency In thermal infrared systems, the ratio of the imaging cone solid angle onto the sensor to the solid angle subtended by the cold aperture (cold shield) within the dewar assembly.

collimated light Light where the rays are all parallel from a given object. A point source at infinity will yield collimated light. This also means that the wavefront is plane.

coma An off-axis aberration where the outer periphery of a lens has a higher or a lower magnification than the central portion of the lens. The resulting image of a point object looks like a small comet.

coma, axial Coma occurring at the center of the field of view introduced by element tilts, decentrations, and/or wedges. This coma generally carries over the field of view.

computer-aided optical design The process of using a lens design computer program to optimize the performance of a lens system and then to evaluate its performance.

computer generated hologram (CGH) Diffractive element which has been calculated by a computer program, often by the mean of an

iterative optimization algorithm. Usually, these elements do not have optical power (not lenses).

concave surface A lens or mirror surface which is inward curving.

conjugate A location which is at an image of another location.

convex surface A lens or mirror surface which is outward curving.

cosmetic effects Those defects in/on a lens or mirror which appear undesirable, but which may have little or no functional impact on performance. Scratches and digs are often classified in this way. Also refers to video anomalies in infrared system imagery.

crown glass One of two main types of glasses, the other being flint glasses. Crown glass is harder than flint and has a lower index of refraction and a lower dispersion.

curvature 1/ radius of a surface.

depolarization Conversion of polarized light to unpolarized light.

depth of field The maximum change in axial position of an object which produces an acceptable image quality. This is typically looser than the Rayleigh criteria, and relates to the optics and the detector. This is a common term in photography.

depth of focus The focus shift corresponding to plus or minus one-quarter of wavefront error. This corresponds to a wavefront error which just meets the Rayleigh criteria and produces imagery which is essentially perfect. The depth of focus is $d = \pm\lambda/(2 \sin^2\theta) = \pm 2 \lambda(f/\text{number})^2$, where θ is the half angle of the final image cone.

dewar A vacuum bottle which is cooled to cryogenic temperatures and holds an infrared detector.

diamond turning A process of ultraprecision machining whereby optical surfaces can be directly produced using a diamond-tipped tool and air bearings, air slides, and, as appropriate, numerical control. Surfaces accurate to within a few tenths of a micrometer are achievable. Nearly all nonferrous metals, as well as several of the infrared transmitting material, can be diamond turned.

diattenuation Dependence of throughput on incident polarization state.

diffraction A spreading of light after a wavefront of light passes by an opaque edge, due to the wave nature of the light or electromagnetic radiation. This spreading causes the formation of the classical Airy disk pattern when a perfect unobscured optical system images a point object.

diffraction efficiency Amount of light diffracted in the desired optical functionality divided by the total amount of light launched onto the diffractive element. Usually the amount of diffracted light in the fundamental diffraction order.

diffractive optics The use of holographic, kinoform, and binary surfaces which use diffraction in controlling wavefronts. Often assists in introducing optical power to a system as well as helping in cancellation of monochromatic and chromatic aberrations.

diffractive optical element (DOE) Diffractive element calculated by computer or by analytic means. Usually these elements have optical power (lenses) in opposition to CGHs.

dispersion The change in refractive index of glass or other refracting materials with color or wavelength.

distortion An aberration which is a change in magnification with field of view. Distortion is typically cubic with field of view and is a mapping error. It does not affect resolution or image quality.

durometer A term indicating the hardness of a material, defined in terms of the material's resistance to indentation under pressure from a standard tool. The durometer scale goes from 0 to 100, with higher values indicating a harder material.

eigenpolarization A polarization state that is unchanged when incident on a component.

electric discharge milling A process in which material is removed from an electrically conducting work piece by the erosive action of repetitive high voltage electric discharges from a moving thin wire or a shaped tool acting as an electrode.

entrance pupil The position along the optical axis of a lens or mirror system where the chief ray would intersect the optical axis if it were not redirected by the lenses or mirrors. Also, the location and size of the image of the aperture stop when looking into the front of the system.

etendue Product of the area of the light beam and the solid angle that the beam includes. Etendue represents the conservation of radiance, and is maintained throughout a given optical system.

exit pupil The position along the optical axis of a lens or mirror system where the chief ray exiting the system appears to have crossed the optical axis. Also, the location and size of the image of the aperture stop when looking into the rear of the system.

f/number The ratio of the focal length to the clear aperture diameter.

First-order optics says that $f/\text{number} = 1/(2 \tan\theta)$, and for an aplanatic system $f/\text{number} = 1/(2 \sin\theta)$, where θ is the half angle of the final image cone.

field lens A lens or group of lenses located at or near an intermediate image which images the exit pupil of an objective lens on the entrance side of the field lens into the entrance pupil of an objective lens on the exit side of the field lens so as to minimize vignetting and increase light throughput.

flint glass One of two main types of glasses, the other being crown glasses. Flint glass is softer than crown, has a higher index of refraction and a higher dispersion.

focal length The distance measured along the optical axis from the image to the plane where the backward-extended axial imaging cone of light intersects the extended input light bundle.

focal plane array (FPA) A linear or two-dimensional matrix of individual detector elements, typically used at the focus of an imaging system.

fringes Dark (and light) bands on an interferogram formed by interference of two light beams. After reflection from a surface with one wave of irregularity, we see two fringes. After transmission through a material with one wave of irregularity on one surface, we see $(n - 1)(1 \text{ fringe})$ or about 0.5 fringe for glass of refractive index $n = 1.5$. A net system error of one fringe of wavefront error is one wave of optical path difference.

galvanic couple Two dissimilar electrical conductors, usually metals, in contact or within an electrolyte develop a difference in electric potential resulting in corrosion of the surface(s) of one or both materials.

gaussian beam A beam of light whose intensity profile is gaussian in cross section. Lasers typically emit gaussian beams.

ghosting An effect in infrared systems where a facet of a polygon mirror adjacent to a facet actually being used is imaging onto a detector.

holographic optical element (HOE) Diffractive element recorded as a hologram in a holographic plate. The object beam can be produced also by either a DOE or a CGH.

interferometer An instrument whereby a test wavefront is caused to interfere with a reference wavefront. Any difference between the reference and the test will show as light and dark fringes, which are typically photographed or viewed using a vidicon, CCD or other sensor.

iterative fourier transform algorithm (IFTA) Iterative algorithm used to optimize CGHs data on a computer. The simplest implementation of an IFTA is the Gerchberg-Saxton algorithm.

kinoform Phase diffractive element fabricated by surface relief modulation.

left-handed image After reflection in a mirror, an image is upside down, and even if it is rotated around the direction of image propagation, it cannot be oriented as the input object.

magnification, lateral The ratio of the image height to the object height. This is also equal to the ratio of the image distance to the object distance, or to the ratio of the image side *f*/number to the object side *f*/number.

magnification, longitudinal The ratio of the image motion along the optical axis to the corresponding object motion along the optical axis. The longitudinal magnification is the square of the lateral magnification.

modulation transfer function (MTF) The ratio of the modulation in the image to the modulation in the object for a sinusoidal object as a function of spatial frequency. The MTF is affected by diffraction and geometrical aberrations. For a perfect system, the maximum resolvable frequency is $1/(\lambda f/\text{number})$ in line pairs per millimeter, and this is where the MTF goes to zero.

Narcissus The effect in an infrared system where the detector “sees” a reflection of itself. If this reflected radiation changes or modulates through scan and/or over a field of view, then less cold radiation is reflected back into the detector off axis and a dark central region or “porthole” appears in the display.

objective lens The primary lens, which takes light from an object and forms an image. An objective lens generally consists of multiple elements in order to minimize aberrations.

optical path difference (OPD) The difference between a reference (or perfect) wavefront and a real wavefront. If the OPD is one-quarter of the wavelength, then the system just meets the *Rayleigh criteria* and the system will be essentially diffraction limited.

outgassing A process of releasing adsorbed or occluded gasses or water vapor from a material. This typically occurs under a vacuum and/or at high temperature.

paraxial The region where the angles between the rays and the optical axis are small, and the approximation that the sines and tangents

of angles can be represented by their value, in radians, is valid. This makes computations fast and easy, and provides for a convenient means of locating, for example, the image position without regard to aberrations.

partial dispersion Difference in index of refraction for two wavelengths. Main dispersion is $(n_F - n_C)$, where $F = 486.13 \text{ nm}$ and $C = 653.27 \text{ nm}$.

physical optics Refers to the field of optical analysis which does not use ray tracing, but rather diffraction phenomenon to predict the behaviour of optical element like lenses, prism, gratings, holograms, diffractive lenses, etc.

power fit to testplate The number of fringes of power seen by the optician when placing in close contact the surface being fabricated and a surface of known radius (the testplate). Each fringe represents one-half wave of sag or difference between the two surfaces.

principal planes The locations within an optical system where incoming collimated light intersects the light directed to the image (if each are extended until they meet).

Rayleigh criteria The rule of thumb developed by Lord Rayleigh that if the difference between the longest and shortest paths leading to a selected focus is less than or equal to one-quarter of the wavelength, then the imagery is nearly indistinguishable from perfect. If a system meets the Rayleigh criteria, the optical path difference is approximately $\frac{1}{4}\lambda$ or less, and the imagery is essentially perfect.

refractive index The ratio of the velocity of the radiation (light) in a vacuum to the velocity of the radiation (light) in a material. The higher the refractive index, the more the radiation "bends" or is refracted at the air-material surface. Radiation incident on a surface obeys Snells law which says $n \sin\theta = n' \sin\theta'$, where n and n' are the refractive indices on each side of the interface and θ and θ' are the angles of incidence and refraction measured from the normal to the surface.

relay lens A lens or lens group which relays a finite object to a remote location at a magnification of unity or some other value.

retardance Optical path difference accumulated between different polarization states.

right-handed image After reflections and refraction in the system, the image is oriented such that if it is rotated around the direction of image propagation, it can be brought to the same orientation as the input object.

sag The height of the curve measured from the chord. Often used to specify an aspheric surface.

scan noise An effect in infrared systems whereby the cooled detector “sees” more radiation or energy (from within the system) at one position in its field of view or scan than another position, therefore producing undesirable cosmetic effects on the system display.

secondary spectrum In the systems where the primary color is corrected, and the blue and red focus brought to the same point, secondary spectrum or secondary color is the distance between the green focus and the red-blue focus.

spherical aberration The axial aberration where rays from the outer periphery of the lens focuses closer (or further) from the lens than the rays closer to the axis. Spherical aberration is typically proportional to the cube of the aperture.

spherochromatism Variation of spherical aberration with wavelength.

stigmatic Perfect imagery in the geometrical sense.

Strehl ratio The ratio of the peak intensity in the diffraction pattern of an aberrated point image to the peak intensity in the diffraction pattern of the same aberration free point image.

vignetting A clipping or truncation of the off-axis ray bundles by elements distant from the aperture stop. Vignetting is usually intentional in visible systems, as elements can be made smaller and lighter in weight while producing better imagery (by eliminating severely aberrated rays). For visible systems, vignetting of 30% to even as much as 50% can typically be tolerated. Vignetting is usually not tolerable in scanning infrared systems.

This page intentionally left blank

BIBLIOGRAPHY

- "A Chart Demonstrating Variation in Acuity with Retinal Position," letter to the editor, *Vision Research*, vol. 14, 1974, p. 589.
- Buralli, D. A. (1991), "Optical Design with Diffractive Lenses," Sinclair Optics, Design Notes 2 (4).
- Glatzel, Erhard. (1980), "New Lenses for Microlithography," 1980 International Lens Design Conference Proceedings, SPIE.
- Holland, L. (1956), *Vacuum Deposition of Thin Films*, London: Chapman & Hall.
- Jacobson, M. (1986), *Deposition and Characterization of Optical Thin Films*, New York: Macmillan.
- Kingslake, R. (1965), *Applied Optics and Optical Engineering*, New York: Academic Press.
- Kingslake, R. (1978), *Lens Design Fundamentals*, New York: Academic Press.
- Kingslake, R. (1983), *Lens Design Fundamentals*, New York: Academic Press.
- Laikin, M. (1991), *Lens Design*, New York: Marcel Dekker.
- Londono, C. (1992), *Design and Fabrication of Surface Relief Diffraction Optical Elements, or Kinoforms, with Examples for Optical Athermalization*, Somerville, MA: Tufts University.
- Lotmar, W. (1971), "Theoretical Eye Model with Aspherics," *Journal of the Optical Society of America*, 61:1522–1529.
- Macleod, H. A. (1986), *Thin-Film Optical Filters*, New York: Macmillan.
- Melzer and Moffitt (1997), *Head Mounted Displays*, New York: McGraw-Hill.
- Morris, G. M., and Dale A. B. (1992), "Diffractive and Binary Optics," *OSA Short Course*.
- Mouroulis, Pantazis (1999), *Visual Instrumentation*, New York: McGraw-Hill.
- Optical Society of America. (1995), *Handbook of Optics*, vols. 1 and 2, New York: McGraw-Hill.
- ORA Seminar Notes. (April 1999), "Design of Efficient Illumination Systems," Pasadena, CA: ORA.
- O'Shea, D. C. (1985), *Elements of Modern Optical Design*, New York: John Wiley.
- The Photonics Directory*, Laurin Publishing Co. (annual).

- Pulker, H. K. (1984), *Coatings on Glass*, Amsterdam: Elsevier.
- Rancourt, J. D. (1987), *Optical Thin Films Users' Handbook*, New York: McGraw-Hill.
- Ray, S. F. *Applied Photographic Optics, Lenses and Optical Systems for Photography, Film, Video and Electronic Imaging*, 2d ed., London, England, Focal Press.
- Rhyins, R. (June 1974), *Laser Focus*, p. 55.
- Riedel, Max J. (2001), "Optical Design Fundamentals for Infrared Systems" tutorial texts in optical engineering; v.TT20, SPIE.
- Shannon, R. (1980), "Aspheric Surfaces," *Applied Optics and Optical Engineering*, R. Kingslake, ed., vol. 8, New York: Academic Press.
- Shannon, R., and James C. W. (1980), *Applied Optics and Optical Engineering*, R. Kingslake, ed., vol. 8, New York: Academic Press.
- Siegman, A. E. (1986), *Lasers*, University Science Books.
- Smith, W. J. (1990), *Modern Optical Engineering*, New York: McGraw-Hill.
- Smith, W. J. (1992), *Modern Lens Design*, New York: McGraw-Hill.
- Smith, W. J. (1997), *Practical Optical System Layout*, New York: McGraw-Hill.
- Stover, J. C. (1990), *Optical Scattering Measurements and Analysis*, New York: McGraw-Hill.
- Strong, J. (1958), *Concepts of Classical Optics*, New York: Freeman.
- Thelen, A. (1989), *Design of Optical Interference Coatings*, New York: McGraw-Hill.
- U.S. Military Handbook for Optical Design* (1987), republished by Sinclair Optics, Fairport, NY.
- Walker, B. (1995), *Optical Engineering Fundamentals*, New York: McGraw-Hill.
- Wetherell, W. (1980), "The Calculation of Image Quality," *Applied Optics and Optical Engineering*, R. Kingslake, ed., vol. 8, New York: Academic Press.
- Williams, C. S. (1989), *Introduction to the Optical Transfer Function*, New York: John Wiley & Sons.
- Yoder, P. (1986), *Opto-Mechanical System Design*, New York: Marcel Dekker.



INDEX

This page intentionally left blank

- 1/f noise (read-out noise), 671
- Abbe, Ernst, 99
- Abbe illumination, 322–323
- Abbe number (V), 99, 103, 629–632, 775
- Abbe sine condition, 39
- aberrations, 59–60, 62, 775
- astigmatism, 75–78, 187
 - axial color, 89–90
 - chromatic, 60
 - coma, 72–75
 - defined, 36–40
 - distortion, 85–89
 - field curvature, 78–80
 - field lenses, 80–85
 - geometrical, 35–40
 - lateral color, 90–91
 - orders of, 55
 - overview, 59–60
 - plane parallel plates, 91–94
 - reduction of with IR materials, 236–239
 - spherical, 60–72, 91, 105, 115, 187
- abrasion, 391, 482
- abscissa, 170
- absorptance, 570
- acceptable quality level (AQL), 500
- achromatic doublet, 89, 97, 134, 588–589, 605–610, 649
- achromatic retarders, 549–552
- acrylic, 111
- active pixel sensors (APS), 667
- ADC (analog to digital converter), 672
- adhesive bond, 456
- adjusting parameter (compensator), 355, 364–365, 373
- aerial image modulation (AIM) curve, 163
- afocal lens, 15, 775
- aft, 136
- AIM (aerial image modulation) curve, 163
- Air Force target, 333–337
- Airy disk, 35–36, 43–45, 91–92, 205, 378, 775, 767
- aliasing, 678–680, 685
- Amici prisms, 153–155
- amplitude gratings, 297–298
- AMTIR materials, 234
- analog to digital converter (ADC), 672
- analytic type diffractives, 263
- analyzers, 517–518
- anamorphic asphere, 116
- angle of deviation, 25
- angle of incidence (AOI), 75, 584
- angle to area, 325
- anomalous dispersion, 104
- antialiasing filter (OLPF), 685–689
- anticounterfeiting, 313
- antireflection (AR) coatings, 275, 497, 571, 705
- AOI (angle of incidence), 75, 584
- aperture stop 11, 20, 29–30, 136, 242–243, 775
- apertures, insufficient sampling of, 748–749
- aplanatic lens, 775
- apoachromatic lens, 775
- apparent field of view, 776
- apparent thickness, 147
- APS (active pixel sensors), 667
- AQL (acceptable quality level), 500
- AR (antireflection) coatings, 275, 497, 571, 705
- area-solid angle product, 325
- arsenic trisulfide, 235
- aspect ratio, 499
- aspheric Fresnel lens, 652
- aspheric lenses, 121–122, 266, 495–496
- aspheric surfaces, 62, 145, 776
- conic surfaces, 117–118
 - defined, 115–117
 - in reflective and refractive systems, 119–124
- specification of, 126–127
- usage guidelines, 124–126

- aspherical diffractive lens, 266
 assembly
 drop-in, 441–444
 lathe, 444–446
 poker chip, 449–451
 astigmatism, 75–78, 92–93, 776
 asymmetrical errors, 359
 athermal singlets, 281
 athermalization, 146, 232, 600
 IR systems, 246–249, 254
 specifications, 746
 techniques, 467–475
 using teflon, 746
 availability, glass, 107
 axial chromatic aberration, 89
 axial color, 89–90, 776
 axial coma, 365, 777
 axial gradients, 7
 axial height, 494
 axial preload, 410, 414–415, 434–438

 back focal distance, 776
 baffles
 design, 718–722
 optima geometry, 722–725
 profile design, 725–726
 thread profiles, 725–726
 barium borate (BBO), 543
 barium fluoride, 235
 barrel distortion, 86
 Bayer pattern, 691–694
 BB (broadband) antireflection
 coating, 571
 BBO (barium borate), 543
 beam diameters, 136, 383–384
 beam direction, 561–562
 beam intensity, 560–561
 beam sampling, 274
 beam separator, 541
 beam shaping, 274, 296
 beam splitting, 274
 beam waist and divergence, 201–203
 beam wander, 239, 242–243
 beamsplitters
 cube, 538–540, 758–759
 dichroic, 574
 polarizing/nonpolarizing, 533–534, 577

 bending, 422, 439–441
 Bessel functions, 40
 bevels, 428–429, 496
 biconic surface, 116
 Bi-directional Reflectance
 Distribution Function (BRDF)
 profiles, 729
 bidirectional scan, 223
 binary optics, 262, 776
 binocular lens design, 642–646
 binocular optics, 156–158
 biomedical applications, diffractive
 optics, 314
 biomorphic sensors, 696
 birefringence, 393, 395, 481, 511
 polarization, 526–528
 stress, 528–530
 birefringent material, 569, 686–687
 birefringent polarizers, 542–543
 bit depth, 672–673
 blackbodies, 531
 blackening edges, 500
 blazed grating, 260–262
 blocking quantities, 497
 blooming, 667
 Boltmann's constant, 671–672
 bond, adhesive, 456–459
 bonded mountings
 mirrors, 459–462
 prisms, 456–459
 boresight error, 25, 776
 bounces
 infrared system, 706–708
 Narcissus, 708–709
 visible systems, 704–705
 Bowers Schmidt telescope, 142
 Bragg, 275, 312
 BRDF (Bi-directional Reflectance
 Distribution Function) profiles, 729–731
 break edges, 496
 Brewster angle, 524
 bright field, 731–736
 brightness, 322
 broadband (BB) antireflection
 coating, 571
 broadband, diffractive optics, 268–269

- calcium fluoride, 235
Camera Link interface, 699
cameras
 connectivity, 697–701
 lenses, 35, 137, 632–640
 single use, with diffractive
 achromat, 755–756
Cassegrain telescope, 18, 119, 124,
 139–140, 192, 194, 597–600, 776
catadioptric lens assemblies, 446–449
catadioptric system, 776
CCD (charged-coupled device), 2, 311,
 660, 664–666
cemented doublets, 501–502
cemented triplet, 757–758
center of gravity (CG), 416
centering tolerance, 362, 492–493
central obscuration, 144, 196, 496
CG (center of gravity), 416
CGHs (computer generated
 holograms), 262, 266, 271–274, 503
chamfers, 496
charged-coupled device (CCD), 2, 311,
 660, 664–666
chemical properties
 material, 483–484
 selecting glass, 108
chemical vapor deposition (CVD), 233
chief ray, 11, 29, 72–73, 182, 776
chromatic aberrations, 60, 104, 144,
 761, 776
circular polarizers, 540–542
“classical” solution, 67
cleanliness, 716–717
clear aperture, 4, 8, 355, 493–494, 777
climate resistance (CR), 484
clipping, 777
CMOS (complementary metal oxide
 semiconductor), 2, 663, 666–668
CNC (computer numerically
 controlled) lens, 488–492
CNC (computer numerically
 controlled) machines, 485
coatings
 optical thin films, 570–577,
 582–583
 polarization of, 525
coefficient of expansion, 145
coefficient of thermal expansion
 (CTE), 394, 423, 436, 470, 758
coherent light, 199
cold finger, 218, 777
cold mirrors, 574
cold shields, 217–219, 706, 777
cold stops, 777
 dewar, cold shields, and, 217–222
 efficiency of, 219–222
collecting optics, 217
collimated light, 272, 777
collimation, laser beam, 203–204
color fringing, 27, 90
color sensors
 Bayer pattern detectors, 691–694
 three color sensor, 694–696
coma, 72–75, 94, 183, 777
compensators, 363
complementary metal oxide
 semiconductor (CMOS), 2, 663,
 666–668
compliance, 424, 456
compound parabolic concentrator
 (CPC), 328
compression molding, 60, 110
compressive stress, 435
computer generated holograms
 (CGHs), 262, 266, 271–274
computer numerically controlled
 (CNC) lens, 488–492
computer numerically controlled
 (CNC) machines, 485
computer performance evaluation.
 See performance evaluation
computer-aided optical design, 777
computer-generated holographic
 (CGH), 503
computing, optical, 311
concave surface, 777
concentric lenses, 497–499
configuration, 2, 3
 hybrid, 142–143
 systems, 138
conformal domes, 433
conic constant, 116
conic surfaces, 117–118
conical interface, 426
conjugate, 777

- Conrady formula, 100
 constraints, 7, 173
 consumer electronics applications,
 diffractive optics, 314–315
 contamination, 390–391
 continuous flanges, 421–422
 contrast, 192, 716
 contrast ratio, 192
 conventional centering, lens
 manufacturing, 488
 conventional lens, 486
 convex surface, 777
 Cooke triplet, 16, 29, 32, 80, 129,
 134–135
 correlation, wavefront or surface,
 376–377
 corrosion, 390–391
 cosmetic effects, 777
 cosmetic tolerances, 496–497
 cost, manufacturing, 502
 Coude, 765
 CPC (compound parabolic
 concentrator), 328
 CR (climate resistance), 484
 critical illumination, 322
 crown glass, 89, 778
 crown materials, 98
 crowned lens rims, 415–416
 CTE (coefficient of thermal
 expansion), 394, 423, 436, 470, 758
 cube beamsplitters, 758–759
 curvature, 115, 116, 778
 cutoff, frequency, 192
 CVD (chemical vapor deposition), 233
 cyclic vibration, 398
- damping, 398
 dark current noise, 670–671
 dark field, 731–736
 data storage, diffractive optics, 311
 DBCs (double bounce combinations),
 713
 decentration, 408, 409, 412, 424, 462
 degree of polarization (DOP), 510,
 519, 577
 demosaicing, 692–694
 Dense Wavelength Division
 Multiplexing (DWDM), 265
- density (ρ), 393
 depolarization, 530–531
 deposition, 578
 depth of field, 58, 635, 778
 depth of focus, 56–58, 778
 design forms, 129
 mirrors and prisms, 147–155
 overview, 129–130
 reflective systems, 138–146
 refractive systems, 131–137
 of visual systems, 155–166
 designers, approach to optical design,
 171–176
 despace, 408, 424, 470
 detector arrays
 CCD detectors, charge read-out,
 664–666
 CMOS, 666–668
 descriptions of, 663–664
 detector response, 669
 sensor array frame integration
 temporal considerations,
 668–669
 detector noise, 669–672
 detector Nyquist frequency, 676–678
 detector phase, 680–682
 detector response, 669
 detector's eye view, 240
 detetector phase, 680–682
 dewar, 217–219, 707, 778
 DFT (Discrete Fourier transforms),
 281
 diametral thermal gradients, 7
 diamond ruling/turning, 291
 diamond turning, 60, 278, 291, 778
 diattenuation, 525–526
 dichroic beamsplitter, 574
 dichroic sheet polarizers, 535
 dielectric mirrors, 570, 574
 dielectrics, 570, 574, 581, 582
 diffraction, 35, 40, 42, 778
 defined, 40–43, 259–262
 efficiency, 278, 282, 287, 292,
 297–301, 304–305
 gratings, 265–266
 orders, 266
 diffraction orders, 266, 274
 diffraction-limited optics, 43–45

- diffraction-limited performance, 53
diffractive optical elements (DOEs),
 262, 266–274, 313, 503
diffractive optics, 778. *See also* optical
 microlithography
 anticounteरting, 313
 biomedical applications, 314
 consumer electronics applications,
 314–315
 data storage, 311
 design and modeling tools, 277–281
 diffraction efficiency, 760–761
 fabrication of, 287–291, 306–308
 imaging applications, 310
 industrial optical sensors
 applications, 313–314
 laser material processing, 313
 market analysis and future
 applications, 316–318
 niche markets, 316
 optical computing and
 interconnections, 311
 optical telecom, 312
 overview, 259–262
 projection display applications,
 315–316
 versus refractive lenses, 761–763
 rules of thumb, 772–773
 spectroscopic applications, 308–310
 types of, 262–277
diffractive surfaces, 769
digital cameras. *See* cameras
digital optics, 262
direct write techniques, 301
Discrete Fourier transforms (DFT),
 281
dispersion, 60, 89, 99–102, 778
distortion, 778
 in 1:1 imaging lens, 739–740
 correcting, 697
 geometrical aberrations, 85–89
 negative or barrel, 87
 positive or pincushion, 87
 sign of, 742–744
distribution, 199
diverger sphere, 341
DOEs (diffractive optical elements),
 262, 266–274, 313, 503
domes, 417, 422, 429–434
DOP (degree of polarization), 510,
 519, 577
double bounce, 704
double bounce combinations (DBCs),
 713
double Gauss lens, 136, 655–657
 element decentration, 626–627
 element wedge, 627–628
 overview, 610–624
 power fit to testplate tolerance,
 624–625
 refractive index/Abbe number,
 629–632
 surface irregularity, 628–629
 thickness, 625–626
doublets, cemented, 501–502
drop-in lens assemblies, 441–444
DWDM (Dense Wavelength Division
 Multiplexing), 265
dynamic diffractives, 265
dynamic sealing, 444
échelette grating, 265
edge blackening, 500
edge filter, 574
edge thickness difference (ETD),
 360, 492
edge thickness value, 608
EDM (electric discharge milling)
 process, 462
effective focal length, 5, 608
effective medium theory (EMT),
 276, 281
efficiency reductions, 761
E-field components, 686
elastomer, 432, 433, 449, 455
elastomeric mountings,
 422–424
electric discharge milling (EDM)
 process, 462
electronic correction, 696–697
element decentration, 363
elements
 in backward, 747–748
 decentration, 626–627
 polarization, 516–517
 wedge, 627–628

- embossing, 293, 313
 EMT (effective medium theory),
 276, 281
 encircled (ensquared) energy, 8, 189–191
 endoscope, 82
 energy, encircled, 189–191
 ensquared (encircled) energy, 8,
 189–191
 entrance pupil, 8, 29, 31, 778–779
 environment, 389–392, 393, 398
 environmental testing methods, 393
 EPROM (erasable programmable read
 only memory), 474
 equivalent focal length, 5
 equivalent refracting surface, 13
 erasable programmable read only
 memory (EPROM), 474
 erosion, 391, 432
 error function (merit function),
 169, 174
 ETD (edge thickness difference),
 360, 492
 etendue, 21, 324–328, 779
 evaluation, performance. *See*
 performance evaluation
 exit pupil, 10, 24, 31, 779
 external ghosting, 227, 245
 extinction ratio, 532, 533, 536–537,
 539, 577
 extraordinary beam, 527
 eyepiece, 138
 eyes
 optics of, 155
 parameters of, 155–158
 relief, 161
 resolution, 163
 sensitivity, photopic, 95
- f/number (f/#), 11, 799
 f/number at used conjugate, 9
 F-θ lenses, 211
 fabrication, diffractive optics. *See*
 diffractive optics
 Fabry-Perot filter, 575
 factors of merit (FOM), 394
 fast axis, 209
 fast Fourier transform (FFT)
 algorithm, 279
 fast Fourier transform (FFT)-based
 numerical propagators, 281
 FDTD (finite difference time
 domain), 281
 FEA (finite element analysis),
 392, 424
 FFT (fast Fourier transform)
 algorithm, 279
 FFT (fast Fourier transform)-based
 numerical propagators, 281
 field curvature, 78–80
 field flattener, 80
 field lenses, 78, 80–85, 779
 fields of view, 1
 horizontal, 5
 insufficient sampling of, 748–749
 final wafer dicing, 306
 finite conjugate, 9
 finite difference time domain
 (FDTD), 281
 finite element analysis (FEA), 392,
 424
 first principal plane, 13
 first-order optics, 11
 first-order parameters, 167
 first-order relationships
 magnification, 17–19, 21–24
 optical invariants, 20–21
 optical power, 19–20
 overview, 15–17
 plane parallel plates, 24–27
 fixed pattern noise (FPN), 672
 Fizeau interferometer, 342–343
 flange, 421–422, 440, 459
 flare, 705
 flat bevel, 429
 flat-field lithography lens, 84
 flexure, 424–425, 462
 flexure mountings, 424–425, 462
 flint, 98
 flint glass, 779
 FLIR (forward looking infrared),
 45, 472
 flux, 322
 focal length, 5, 8, 13, 779
 focal plane array (FPA), 217, 218,
 222, 779
 focusing lens (IR), 217

- focusing, of gaussian beams, 204–205
FOM (factors of merit), 394
force (F), 393
forward looking infrared (FLIR),
 45, 472
Fourier approximations, 279
Fourier CGH, 272
FPA (focal plane array), 217, 218,
 222, 779
FPN (fixed pattern noise), 672
F-θ lenses, in laser scanners, 211–212
FR (resistance to staining), 484
frequency, fundamental (natural),
 392, 398–400
Fresnel approximations, 279
Fresnel CGHs, 272
Fresnel equations, 267, 279, 400, 704
Fresnel lenses, 266, 268, 275, 278,
 305–306, 651
Fresnel reflection, 704, 706, 708
Fresnel zone plate, 266, 268
fringes, 779
full field of view, 5
full frame (snap-shot) integration,
 668–669
fundamental frequency, 398–400,
 461
fungus, 391–392
- Galilean telescope, 137
gas lasers, 208
gaussian beam imagery
 application of in laser systems,
 208–210
 beam waist and divergence,
 201–203
 collimation of laser beams,
 203–204
 F-θ lenses in laser scanners,
 211–212
 overview, 199–201
 propagation and focusing of,
 204–205
 truncation of, 205–208
Gaussian intensity distribution, 200
Gauss-type lens, 83
GDS2 layouts, 293
generators, polarization, 517
- geometrical aberrations
 astigmatism, 75–78
 axial color, 89–90
 coma, 72–75
 defined, 36–40
 distortion, 85–89
 field curvature, 78–80
 field lenses, 80–85
 lateral color, 90–91
 overview, 59–60
 plane parallel plates, 91–94
 reduction of with IR materials,
 236–239
 spherical, 60–72
geometrical optics, 295
geometrical scan effects, 242–243
Gerchberg-Saxton (G-S) algorithm, 273
germanium, 229–233
ghosting, 227, 239, 245, 779
ghosts
 infrared system considerations,
 706–708
 Narcissus, 708–709
 stray light, 713
 visible systems, 704–705
Gigabit Ethernet (GigE) Vision,
 700–701
Glan-Thompson polarizer, 542
glass
 anomalous dispersion, 91
 crown, 89
 flint, 89
 material properties, 95–96
 material specifications, 480–485
 parametric examples of, 102–106
 plastic optical materials, 109–111
 selecting, 103, 106–113
 wrong choice, 747
 wrong type in precision lens
 system, 754–755
glass map, 96–102
glass material, 485
glasses
 availability, 107
 chemical properties, 108
 normal, 107
 thermal properties, 109
transmittance, 107–108

- Glatzel lens, 256, 620
 global minimum, 170
 Gradium axial refractive index, 653
 grating equation, 260, 277
 gray scale intensity mapping, 303
 gray scale optical lithography, 301–303
 Gregorian design form, 140–141
 Gregorian telescope, 140–141
 grinding, in lens manufacturing, 486–488
 G-S (Gerchberg-Saxton) algorithm, 273
 half-wave optical thickness (HWOT) layer, 570
 Hammer optimization, 615
 hardware design
 off-the-shelf optics, 587–594
 optomechanical design, 600–602
 pupil matching, 594–595
 stray light control, 596–600
 harmonic diffractive lenses, 269
 Hartman test, 345–346
 Hartmann formula, 100
 HEBS (High Energy Beam Sensitive - glass), 301
 Helmholtz invariant (Lagrange invariant), 20
 Helmholtz's wave equation, 278
 hemispheres, 499
 Herschel condition, 38
 High Energy Beam Sensitive - glass (HEBS), 301
 high numerical aperture (NA) systems, 534
 high-energy radiation, 392
 HIP'ed (hot isostatic pressed), 233
 HOEs (holographic optical elements), 262, 263, 276–277
 holographic diffusers, 264
 holographic exposure, 287–290
 holographic grating, 276–277
 holographic optical elements (HOEs), 262, 263, 276–277
 holographic recording, 276
 homogeneity, 481
 hoop stress, 439
 horizontal field of view, 5
 hot isostatic pressed (HIP'ed), 233
 hot mirrors, 574
 housing, 175
 Hubble telescope, 470, 750–754
 human eyes. *See* eyes
 humidity, 390–391, 431
 Huyghen's principle, 278
 HWOT (half-wave optical thickness) layer, 570
 hybrid achromat, 310
 hybrid design, 173
 hybrid optical compound lenses, 271
 hybrid optical elements, 270–271
 hybrid optics, 270–271, 282–287
 hybrid systems, 142–143
 hyperhemispheres, 499
 IAD (ion-assisted deposition), 578, 585
 IBD (ion-beam deposition), 585
 IC (integrated circuit) industry, 291
 ID (inside diameter), 409
 IEEE1394b, 699–700
 IFTA (iterative Fourier transform algorithm), 273
 illumination systems
 Abbe, 322–324
 etendue, 325–328
 Köhler, 323–324
 optical invariant, 324–325
 other types of, 329–331
 overview, 321–322
 images
 anomalies, 239–246
 degradation of, 356–359
 formation of, 660–663
 orientation of, 756–757
 quality of, 8, 35–47
 upside down/rotated, 749–750
 imaging. *See also* thermal infrared imaging
 applications, diffractive optics, 310
 two-dimensional array, 663
 imaging optical system, 1–4
 increment, 168
 industrial optical sensors
 applications, diffractive optics, 313–314

- infrared (IR) materials
AMTIR materials, 234
arsenic trisulfide, 235
germanium, 229–233
magnesium fluoride, 234
overview, 229
reduced aberrations with,
 236–239
sapphire, 234–235
silicon, 233
zinc selenide, 234
zinc sulfide, 233
infinity corrected objectives, 9
infrared spectral regions, 96
infrared detectors, 222
infrared imaging. *See* thermal
 infrared imaging
injection molding, 60, 110, 284, 307
inside diameter (ID), 409
integrated circuit (IC) industry, 291
integrating spheres, 534
integration time, 668
intensity, 322
interconnections, optical, 311
interfaces, 771
 on bevels, 428–429
 conical (tangential), 426–427
 kinematic, 409
 rim contact, 410, 444
 sharp corner, 426, 435, 442
 spherical, 428
 surface contact, 426–429
 toroidal, 427, 435–436
interference, 569, 574
interferogram-type, diffractive optics,
 269–270
interferometry, 340–344, 780
internal ghosting, 227, 245
ion beam etching, 302
ion-assisted deposition (IAD), 578,
 585
ion-beam deposition (IBD), 585
IR. *See* infrared
iterative Fourier transform algorithm
 (IFTA), 273

Johnson Noise, 671–672
Jones Calculus, 519–522

Jones matrix, 516–517
Jones vector, 516–517, 518

kinematic interface, 409
kinoform period, 651
kinoforms, 262, 780
Knoop hardness, 482
Köhler illumination, 323–324

LAF (light attenuating film), 302
Lagrange invariant (Helmholtz
 invariant), 20
landscape lens, 131–133
laser beam
 collimation, 203–204
 truncation, 206, 207
 waist, 201–203, 206
laser diodes, 209
laser material processing, 313
laser scanners, F-θ lenses in, 211–212
laser systems, gaussian beams in,
 208–210
lateral color or color fringing, 90–91,
 94, 697, 780
lateral magnification, 17, 780
lateral preload requirements, 414–415
lathe assembly process, 444
lathe lens assemblies, 444–446
LCD (liquid crystal display), 512
LEDs (light-emitting diodes), 321
left-handed image, 780
lenses
 achromatic doublet, 134, 588–589
 aspheric, 495–496
 assemblies, 441–449
 bending, 62
 CNC, 488–492
 conventional, 486
 Cooke triplet, 134–135
 design, optimization case studies
 achromatic doublet, 605–610
 binoculars, 642–646
 digital cameras, 632–642
 double Gauss, 655–657
 double Gauss lens design, 610–632
 error function construction,
 603–605
 simple lenses, 646–655

- lenses (*Cont.*)**
 - diffraction, 267–268, 761–763
 - double Gauss, 136
 - doublets, 590–591
 - fabrication, 445, 483, 489
 - hood, 728–729
 - landscape, 131–133
 - mounting techniques, 419–425
 - optimizing systems, 168–171
 - performance, 183
 - Petzval, 136
 - rim, 415–416, 444
 - sample design problem, 176–178
 - simple, 646–655
 - single-element, 131, 588–589
 - singlets, 590–591
 - telephoto, 136–137
 - unnecessary elements, 744
 - wide-angle, 137
 - Zeiss Tessar, 135
- lenslet, 330
- LensView, 172
- LGA (local grating approximation), 277
- light
 - insufficient, 745–746
 - polarized, 508–510, 513–523
 - stray, 703–715, 736–737
- light attenuating film (LAF), 302
- light pipes, 326–327
- light-emitting diodes (LEDs), 321
- light-gathering power, 325
- lightweight mirror structures, 404–405
- linear array scanning, 662
- linear diffraction gratings, 265
- linespread function (LSF), 338
- liquid crystal display (LCD), 512
- lithium fluoride, 235
- local grating approximation (LGA), 277
- local minimum, 171
- long wave IR (LWIR), 706
- longitudinal magnification, 18, 780
- longitudinal spherical aberration, 61
- long-wave infrared (LWIR), 45, 213, 706
- Lotmar eye model, 161
- low light level systems, 711
- lower marginal ray, 182
- LSF (linespread function), 338
- LWIR (long-wave infrared), 45, 213, 706
- Lyot filter, 564
- Lyot stop, 144, 600
- machine vision optics, 596–597
- MacNeille cube, 538
- magnesium fluoride, 234
- magnetorheological finishing (MRF), 490
- magnification, 21–24
 - lateral, 17–18
 - longitudinal, 18–19
 - visual system, 21
- main dispersion, 99, 101
- Maksutov telescope, 141
- maltese cross, polarization, 531–533
- Mandler, Walter, 136
- Mangin mirror, 763–765
- manufacturing
 - antireflection coatings, 497
 - aspect ratio, 499
 - aspheric lens, 495–496
 - bevels, chamfers, and break edges, 496
 - blocking quantities, 497
 - cemented doublets, 501–502
 - centering tolerance, 492–493
 - clear aperture, 493–494
 - CNC lens fabrication, 488–492
 - component testing, 500
 - concentric lenses, 497–499
 - conventional centering, 488
 - conventional lens fabrication, 486
 - cosmetic tolerances, 496–497
 - edge blackening, 500
 - errors, 347–348
 - fine grinding/polishing, 486–488
 - hemispheres and
 - hyperhemispheres, 499
 - material, 480–485
 - overview, 479–480, 485–492

- manufacturing (*Cont.*)
power tolerance, 495
radius tolerance, 494–495
relative cost, 502
sag tolerance, 494
segmenting, 500
sourcing, 502–504
surface irregularity, 495
thickness tolerance, 494
thin edges, 499–500
manufacturing drawing, 505
marginal ray, 11, 20
mask layouts, 293
material
design, 480–482
fabrication, 482–485
infrared, 229–235
properties, 95–96
MCM (multichip modules), 311
mechanical motion, using
polarization to control,
565
mechanical parameters and
properties
material, 482–483
optomechanical design,
393–396
melt design, 353
MEMs (microelectromechanical
systems), 314
meridional rays, 72
merit function (error function),
169, 174
metal wire polarizers, 535–537
metallic mirror, 574
metal-oxide-semiconductor (MOS),
664
microelectromechanical systems
(MEMs), 314
microlens array, 272
microlenses, 673
microlithographic fabrication,
292
microlithography lenses, 83
microscope, visual, 22
midwave IR (MWIR), 213,
706
Mie scattering, 712
- mirrors. *See also* prisms
cold, 574
design forms, 147–155
hot, 574
lightweight, 404–405
Mangin, miscoated, 763–765
metallic, 462–465
mountings, 459–467
nonmetallic, 465–467
parabolic, 139
primary, 139
secondary, 139
modular design, 401–404
modulation, 191, 304
modulation transfer function (MTF),
8, 179, 181, 191–198, 333, 592, 780
optical quality factor, 379–383
optical testing, 337–340
Monte Carlo analysis, 631
Monte Carlo tolerancing, 357
MOS (metal-oxide-semiconductor),
664
Mouroulis, 163
MRF (magnetorheological finishing),
490
MTF. *See* modulation transfer
function
Mueller calculus, 519–522
Mueller matrix, 520
multiaperture window, 432
multichip modules (MCM), 311
multilevel Fresnel lens efficiency,
305–306
multilevel optical microlithography
diffraction efficiency calculations,
297–301
example of fabrication, 294–295
successive mask alignments, 296–297
multiple block, 487
multiwave retarders, 547–549
mutifocus lens, 272
MWIR (med-wave IR),
213, 706
- NA (numerical aperture), 8–10, 267,
732
- Narcissus effect, 240–242, 708–709
- narrowband interference filter, 575

- near infra-red (NIR), 214, 234, 669
 negative distortion, 86
 Newton, Isaac, 96
 Newtonian telescope, 139, 153
 Newton's equation, 16
 NIR (near infrared), 214, 234, 669
 Nocitlux, 136
 nodal points, 14
 noise
 aliasing, 678–680
 bit depth, 672–673
 dark current, 670–671
 detector phase, 680–682
 fixed pattern noise (FPN), 672
 improving pixel fill factor
 limitations, 673–675
 lens limited versus detector limited
 systems, 689–690
 OLPF (antialiasing filter), 685–689
 overview, 669–670
 pixel pitch/detector Nyquist
 frequency, 676–678
 pixel sizes/sampling intervals,
 682–685
 quantization, 672
 quantum (shot), 671
 read-out ($1/f$), 671
 reset noise due to thermal noise
 (Johnson Noise), 671–672
 standard sensor sizes, 675–676
 system level MTF, 690–691
 nonkinematic prism mounting, 410
 nonuniformity correction (NUC),
 708
 notch filters, 575
 NUC (nonuniformity correction),
 708
 null lens, 341
 numeric type diffractives, 264
 numerical aperture (NA), 8–10, 267,
 732
 numerical propagators, 281
 Nyquist frequency, 635, 676–678, 693,
 768
 objective lens, 780
 OD (outside diameter), 409
 Offner design, 257
 off-the-shelf components, 34
 off-the-shelf optics, 587–591
 development of lab mockup, 595
 example, 591–594
 precision lens assemblies, 587–588
 single elements/achromatic
 doublets, 588–589
 singlets/doublets, 590–591
 working effectively with, 589–590
 OLPF (optical low pass filter),
 685–689
 OPD. *See* optical path difference
 operands, 605
 operational environments, 389
 operational specifications, 7
 OPL (optical path length), 38
 optical computing and
 interconnections, 311
 optical design
 designer approach to, 171–176
 general topics, 767–770
 optimizing lens systems, 168–171
 process, 167–168, 172
 sample problem, 176–178
 optical element prints, 175
 optical invariant, 20, 324–325
 optical low pass filter (OLPF),
 685–689
 optical microlithography
 diffraction efficiency for industrial
 applications, 304–305
 final wafer dicing, 306
 gray scale, 301–303
 multilevel, 294–301
 multilevel Fresnel lens efficiency,
 305–306
 overview, 291–293
 techniques, 303–304
 optical path difference (OPD), 70, 357,
 364, 374, 393, 780
 depth of focus, 56–58
 peak-to-valley and rms wavefront
 error, 52–54
 performance evaluation, 189
 and Rayleigh criteria, 49–52
 wave aberration polynomial, 55–56
 optical path length (OPL), 38
 optical performance, 169, 356

- optical pickup (OPU) lens, 270
optical power, 19, 63–67
optical quality factor (OQF),
 379–383
optical security, 317
optical systems
 basic parameters, 4–11
 first-order relationships, 17–27
 overview, 1–4
 polarization analysis, 555–559
 starting, 168, 172
 terminology, 11–15
 for UV, 255–258
optical telecom, 312
optical testing
 interferometry, 340–344
 modulation transfer function
 (MTF), 337–340
 other tests, 344–346
 overview, 333
 with standard 1951 U.S. Air Force
 target, 333–337
optical thin films
 coatings, 570–584
 overview, 569–570
 production cost, tolerances, quality,
 584–585
optical variable devices (OVDs), 313
optic-to-mount interface, 408–414
optimization, 168–171, 174, 272
optomechanical design, 600–602
 applicable guidelines, 393
 athermalization techniques,
 467–475
 axial and lateral preload, 414–415,
 434–438
 bending effects in rotationally
 symmetric optics, 439–441
 environmental considerations,
 389–392
 environmental testing methods,
 393
 hardware design, 600–602
 incorporating prisms, 452–459
 individual lens mounting
 techniques, 419–425
 interfaces for other optical
 components, 416–419
- optomechanical design (*Cont.*)
 mechanical parameters and
 properties, 393–396
 mirror mountings, 459–467
 modular construction, 401–405
 mounting windows, shells, and
 domes, 429–434
 multiple component lens
 assemblies, 441–452
 overview, 389
 radial stresses, 439
 rigid housing configurations, 400–
 401
 shock, 400

- paraxial optics, 39, 62
 paraxial region, 39
 partial dispersion, 101–102, 781
 patent, 172
 patterned polarizers, 543
 peak-to-valley (P-V) optical path difference (OPD), 52–54
 Pechan prisms, 151–153
 Penta prism, 147, 148
 performance evaluation
 defined, 179–180
 encircled energy, 189–191
 modulation transfer function, 191–198
 optical path difference, 189
 optical testing, 333–346
 ray trace curves, 181–187
 resolution, 180–181
 spot diagrams, 187–189
 periodic vibration, 399
 periscope, 82
 Petzval lens, 80, 136
 Petzval sum, 83
 Petzval surface, 78–79
 phase
 gratings, 298–301
 using polarization to control, 563
 photoelectric effect, 663
 photomask layers, 296
 photomasks, 294
 photometry, 21, 322
 photonic crystals, 274–276
 pincushion distortion, 86
 pitch button, 487
 pixels
 color sensing within, 695–696
 improving fill factor limitations, 673–675
 pitch, 676–678
 sizes, 682–685
 plane of incidence, 149
 plane parallel plates, 24, 91–94
 plastic optical materials, 109–111
 Plössl form, 742
 point-spread function (PSF), 8, 51, 195–196, 387
 Poisson's ratio (ν), 394
 poker chip lens assemblies, 449–452
 polarization
 achromatic retarders, 549–552
 analysis of optical system, 555–559
 behavior/components, 510–512
 birefringent polarizers, 542–543
 circular polarizers, 540–542
 controlling beam direction, 561–562
 controlling mechanical motion, 565
 controlling phase, 563
 controlling spatial distribution of light, 562–563
 controlling system
 transmission/optical beam intensity, 560–561
 controlling transmission wavelength, 564–565
 dichroic sheet polarizers, 535
 light, mathematical description of, 513–523
 metal wire polarizers, 535–537
 minimizing problems in optical design, 559
 overview, 507–508
 patterned polarizers, 543
 phenomena, 525–534
 polarized light, 508–510
 polarizing beamsplitter cubes, 538–540
 retarder mathematics, 546–547
 retarders, 544–549, 552–555
 switchable polarizers, 544
 and telescopes, 765–766
 polarizers, 508, 531–544
 birefringent, 542–543
 circular, 540–542
 dichroic sheet, 535
 metal wire, 535–537
 patterned, 543
 switchable, 544
 polarizing beamsplitter cubes, 538–540
 polarizing beamsplitters, 533–534, 577

- polishing, in lens manufacturing, 486–488
polycarbonate, 111
polystyrene, 111
Porro prism, 642, 643
positive distortion, 86
potassium bromide, 235
power fit to testplate, 624–625, 781
power spectral density (PSD), 399, 400
power tolerance, 495
precision lens assemblies, 587–588
precision potting, 492
preload
 axial, 409–412
 variation with temperature, 468
pressure
 effects of, 390
 optomechanical design, 390
primary aberrations, 135
primary mirror, 19, 138
principal plane, 13, 14, 781
principal ray, 20, 29
prisms, 147–155
 color splitting, 694–695
 design forms, 147–155
 mountings, bonded, 456–459
 in optomechanical design, 452–459
productibility
 adjusting parameters, 364–365
 beam diameter, and surface irregularities, 383–384
 correlation, as relates to performance, 376–377
 effect to MTF, 379–383
 effect to spot diameters, 377–379
 example of tolerance analysis, 367–373
 forms of tolerances, 359–364
 image degradations, 356–359
 overview, 347–348
 results, 384–388
 surface irregularities, 374–376
 testplates, 348–353
 tolerancing optical systems, 353–356
 for various cost models, 366–367
production cost, optical thin films, 584–585
projection display applications, diffractive optics, 315–316
propagation, of gaussian beams, 204–205
propagators, numerical, 281
protective bevels, 428
PSD (power spectral density), 399, 400
PSF (point-spread function), 8, 51, 195–196, 387
pupil-forming optics, 159–160
pupil-forming visual systems, 159–160
pupils, 323
 diameter, 181
 entrance and exit, 31
 matching, 34, 594–595
 problems with, 744–745
pushbroom configuration, 662
pushbroom scanning, 223
quality factor, 276
quantization noise, 672
quantum efficiency, detector response, 669
quantum noise, 671
quarter wave retarder, 511
quarter-wave optical thickness (QWOT) layers, 570
radial stresses, variations with temperature, 439
radial thermal gradients, 7, 193
radiant exitance, 215
radiation
 high energy, 392
 thermal, 394
radiofrequency identification (RFID) devices, 313
radiometry, 21, 322
radius, 115
radius tolerance, 494–495
random vibration, 399
Rayleigh criteria, 49–52, 180, 781
Rayleigh range, 203, 204

- Rayleigh scattering, 712
 Rayleigh-Sommerfeld diffraction formulation, 279
 rays, 507
 - chief or principal, 20
 - marginal, 20, 182
 - sagittal, 184
 - skew, 184
 - tangential, 184
 - trace curves, 181–187
 - tracing, 277–278, 282–287, 556–559
 read-out noise (1/f noise), 671
 rear projection (RPTV) architectures, 315
 reflectance, 572
 reflecting prisms, 147
 reflective optics, 119, 144
 reflective systems
 - design forms, 138–146
 - spherical/aspheric surfaces, 119–124
 refraction, 259–260
 refractive index, 46, 69, 108, 629–632, 781
 refractive optics, 146
 refractive systems
 - design forms, 131–137
 - spherical/aspheric surfaces, 119–124
 region of interest (ROI), 667
 relative illumination, 8
 relative partial dispersion, 101
 relay lens, 781
 reoxidize material, 581
 reset noise due to thermal noise, 671–672
 resistance to acid (SR), 484
 resistance to alkali (AR), 484
 resistance to staining (FR), 484
 resolution, 179, 180–181
 resolution element, 1, 2
 resolving power, 180
 responsivity (R), 669
 retaining ring (retainer), 419–420, 431, 447, 459
 retardance, 526–528, 544–555, 566
 retarders, 511
 - achromatic, 549–552
 - mathematics, 546–547
 - overview, 544–546
 - special, 554–555
 - variable, 552–554
 RFID (radiofrequency identification) devices, 313
 right circularly polarized, 513
 right-handed image, 781
 rigid housing configurations, 400–401
 rigorous electro-magnetic modeling techniques, 281–282
 rim contact interface, 439
 rim contact lens, 444
 Ritchey-Chrétien Cassegrain telescope, 120, 124
 Ritchey-Chrétien telescope, 140
 RMS (root-mean-square) blur diameter, 8
 RMS (root-mean-square) wavefront error, 8, 52–54, 768
 ROI (region of interest), 667
 rolling integration, 668
 room temperature vulcanizing (RTV), 422–424
 root sum squared (RSS), 357, 670
 root-mean-square (RMS) blur diameter, 8
 root-mean-square (RMS) wavefront error, 8, 52–54, 768
 RPTV (rear projection) architectures, 315
 RSS (root sum squared), 357, 670
 RTV (room temperature vulcanizing), 422–424
 Rudolph, Paul, 135
 sag tolerance, 494
 sagittal plane, 73
 sagittal rays, 184
 sample lens design, 176–178
 sapphire, 234–235
 sawtooth grating, 291
 scalar theory, 278, 281
 scalar-diffraction-based tools, 278–281

- scan efficiency, 228
scan noise, 239, 243–245, 781
scan-converting electronics, 223
scanning
 bidirectional, 223
 linear array, 662
 methods, 222–229, 661–662
 polygon mirror, 226–228
 push broom, 223–227
 serial, 223
 single point detector, 661–662
scatter path analysis, stray light, 714–715
scattering
 avoiding stray light, 736–737
 bright field/dark field, 731–736
 cleanliness, 716–717
 Mie, 712
 Rayleigh, 712
 spheres, 534
 suppression techniques, 717–731
 veiling glare, 715–716
scene energy, 219
Schmidt telescope, 120, 141–142
Schott catalogue, 107
Schott, Otto, 99
Schwarzschild objective, 256
Schwarzschild reflective microscope, 256–257
sealing, 417–419
second principal plane, 13
secondary axial color, 187
secondary mirror, 19, 138
secondary spectrum, 101, 102, 105, 782
segmenting, 500
Self, S.A., 204
self-weight deflection, 398–400, 417
semikinematic interface, 409
semikinematic prism mounting, 453
sensor systems
 camera connectivity, 697–701
 camera link, 697–699
 Gigabit Ethernet (GigE) Vision, 700–701
 IEEE1394b, 699–700
 USB, 700
sensor systems (*Cont.*)
 color sensors, 691–696
 detector arrays, 663–669
 electronic correction, 696–697
 image formation, 660–663
 noise, 669–691
 overview, 659–660
serial scanning, 223
Serrurier truss, 407
shading, 239, 246
sharp corner interface, 426
shells, 417, 429–434
shims, 413, 422, 451
shipping environments, 389
shock, 392, 400, 408, 456
Short Wavelength Spectrometer, 402
shortened Schmidt Cassegrain, 141
shot noise, 671
Siegman, A.E, 204
silicon, 233
simple lenses, 646–655
single element lens, 588–589
single lens reflex (SLR) camera lenses, 716
single point detector scanning, 661–662
single point diamond turning (SPDT), 402–403, 449
single-element lens, 131
singlet designs, 648
singlets lens, 590–591
skew rays, 73, 184
slow axis, 209
SLR (single lens reflex) camera lenses, 716
snap-shot (full frame) integration, 668–669
Snell's equation, 260
Snell's law, 40
sodium fluoride, 235
Sol-Gel process, 307
solid-state lasers, 209
Sorbothane, 456
sourcing, 502–504
spacer, 414, 422, 445

- spatial distribution of light, 562–563
spatial frequency, 193, 768
spatial homogenization, 321, 326
SPDT (single point diamond turning), 402–403, 449
specifications, 5
spectral bands, thermal infrared imaging in
athermalization, 246–249
cold stop efficiency, 219–222
dewar, cold stops/shields, 217–219
image anomalies, 239–246
IR materials, 229–235
optical systems for UV, 255–258
overview, 213–217
reduced aberrations with IR materials, 236–239
scanning methods, 222–229
system design examples, 250–255
spectral dispersion, 265, 270
spectral range, 6
spectral resolution, 266
spectral shift, 569
spectroscopic applications, diffractive optics, 308–310
specular baffles, 720
spheres, scattering/integrating, 534
spherical aberration, 35, 104, 187, 782
classical solution, 63–67
optimum solution, 67–72
overview, 60–63
spherical interface, 428
spherical surfaces
conic surfaces, 117–118
defined, 115
in reflective and refractive systems, 119–124
spherocallens rims, 415–416
spherochromatism, 103, 187, 287, 782
splitting optical power, 136
spot diagrams, 181, 187–189
spot diameters, 189, 377–379, 770
spot radius, 189, 770
spot size, 189, 770
spring constraints
mirror mountings, 459
prisms, 453–456
spurious resolution, 197–198
SR (resistance to acid), 484
standard sensor sizes, 675–676
star test, 344
staring or mosaic arrays, 224, 663
static sealing, 417
step bevel, 429, 446
stiffness, 392
stigmatic, 782
stoichiometry, 581
Stokes vectors, 518–519
storage specifications, 7
strain, 394
stray light, 146, 262, 595–600, 703–737
avoiding, 736–737
baffling, 597–600
control, 596–600
ghost analysis, 713
modeling scatter, 715
scatter path analysis, 714–715
scatter sources, 703–711
scatter types, 711–712
Strehl ratio, 782
stress
bending, 439–441
birefringence, 528–530
compressive, 394, 435
defined, 394
hoop, in cell wall, 439
influence of on refractive index, 108
tensile, 394, 434–435
Striae Grade AA (P), 481
structural scatter, 710–711
Strutt, William, 50
subwavelength (SW) diffractives, 264, 274–276
subwavelength gratings (SWG), 264
subwavelength structures, 264, 274–276

- support structure configurations
optic-to-mount interface,
 408–414
 overview, 405–408
surface contact design, 410
surface contact interface shapes,
 426–429
surface irregularities, 374–376, 495,
 628–629
surfaces, uncoated, polarization of,
 523–524
survival environments, 424
SW (subwavelength) diffractives, 264,
 274–276
Sweatt model, 278
SWG (subwavelength gratings), 264
switchable polarizers, 544
 S_y (yield stress), 394
symmetrical errors, 359
symmetry fore, 136
Synthetic Hologram, 263, 313
system specifications, 45–47
system transmission, 8, 560–561
systems analysis, 45
- tangential interface, 426–427
tangential oblique spherical
 aberration, 78
tangential rays, 184
tapered light pipe, 327
Taylor expansion polynomial, 55
Taylor, H. D., 134
TEDY tolerance designation, 370
telecentricity angle, 675
telecom, 312
telephoto lens, 136–137
telephoto ratio, 137
telescopes. *See also brand names of specific telescopes*
 Hubble, null lens problem,
 750–754
 and polarization, 765–766
temperature
 effects on axial preload, 436–438
 gradients, 394, 397
 and optomechanical design,
 389–390
tensile stress, 441
- Tessar lens, 173
test glasses, 349
testplates, 348–353
TETY tolerance designation, 370
TFRN tolerance designation, 370
thermal conductivity (k), 394
thermal defocus, 248–249
thermal diffusivity (D), 394
thermal expansion coefficients (CTE),
 394, 423, 436, 470, 758
thermal gradients, 7
thermal infrared imaging
 athermalization, 246–249
 cold stop efficiency, 219–222
 dewar, cold stops, and cold shields,
 217–219
image anomalies
 geometrical scan effects, 242–243
 ghosting, 245
 Narcissus effect, 240–242
 overview, 239–240
 scan noise, 243–245
 shading, 246
IR materials, 229–239
optical systems for UV, 255–258
overview, 213–217
scanning methods, 222–229
system design examples,
 250–255
- thermal properties
 material, 484–485
 selecting glass, 109
- thermal soak, 7
- thickness tolerances (TTHI), 370,
 625
- thin edges, 499–500
- threaded retaining rings, 419–421
- threads
 baffle design, 718–722
 baffle profile design, 725–726
 optimal baffle geometry,
 722–725
 profiles, 726–728
- three bounce rule, 717–718
- three color sensor
 color sensing within pixels,
 695–696
- color splitting prisms, 694–695

- three-mirror anastigmat (TMA), 143
- throughput, 325
- tilt, 405, 416, 424, 429, 459
- tilted plate, 76
- TIR (total indicator runout), 363
- TIR (total internal reflection), 714
- TIRR tolerance designation, 370
- TIRY tolerance designation, 370
- TMA (three-mirror anastigmat), 143
- tolerances
 - analysis, 175
 - example, 348
 - forms of, 355
 - optical thin films, 584–585
 - typical, 366–367
- tolerancing
 - adjusting parameters, 364–365
 - beam diameter and surface irregularity, 383–384
 - correlation and performance, 376–377
 - effect to MTF, 379–383
 - effect to spot diameter, 377–379
 - example analysis, 367–373
 - forms of tolerances, 359–364
 - image degradations, 356–359
 - of optical system, 353–356
 - optical systems, 353–356
 - overview, 347–348
 - results, 384–388
 - surface irregularities, 374–376
 - testplates, 348–353
 - for various cost models, 366–367
 - “top-hat” intensity profile, 199
- toroid or toric, 116, 117
- toroidal interface, 427–428
- total indicator runout (TIR), 363
- total internal reflection (TIR), 714
- transfer lens, 342
- transmission sphere, 341
- transmission wavelength, 564–565
- transmittance, 8, 107–108, 575
- transverse ray aberration curves, 182–188, 643–644
- trusses
 - Hubble Telescope, 470
 - Serruer, 407–408
- TTHI (thickness tolerances), 370, 625
- tunnel diagram, 147
- two-dimensional array imaging, 663
- Twyman-Green interferometer, 340, 342
- ultraviolet (UV) lasers, 209
- ultraviolet (UV) lens, 256
- ultraviolet (UV) spectral regions, 96
- ultraviolet (UV) systems, 255–258
- uncoated surfaces, polarization of, 523–524
- unobscured aperture systems, 143–144
- unpolarized light, 509–510
- upper marginal ray, 182
- USB, 700
- UV. See ultraviolet
- variable retarders, 552–554
- variables and constraints, 173–174
- variation with temperature preload, 467
- veiling glare, scattering, 715–716
- vibration, 392, 398–400, 424
- vignetting, 32–34, 83, 675, 782
- visual acuity, 161
- visual optics, 155, 161–165
- visual systems design
 - parameters of human eye, 155–158
 - pupil-forming, 159–160
 - requirements for, 161–166
- wave aberration polynomial, 55–56, 67
- wavefronts, 42
- waveguide gratings, 275
- wavelength band, 6
- wavelength weights, 7
- wedge, 24

- whiffletree plate, 465
whisk-broom scanning, 662
wide-angle lens, 137
windows, 429–434
wire grid polarizers, 510
wire lens, 4
working f/ number, 9
- YAG (yttrium aluminum garnet),
 313
yield stress (σ_y), 394
- Young's modulus (E), 394
yttrium aluminum garnet (YAG),
 313
- Zeiss Tessar lens, 135
Zemax, 107
zero order grating, 275, 281
zero order retarders, 547–549
zinc selenide, 234
zinc sulfide, 233
zoom periscope, 740–742